

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:19:50 ; Search time 49.13 Seconds

(without alignments)  
13.691 Million cell updates/sec

Title: US-09-712-819A-1

Perfect score: 32

Sequence: 1 FTLTISS 7

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 455

Minimum DB seq length: 0

Maximum DB seq length: 7

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	34.4	4	2	A40135
2	11	34.4	5	2	PT0644
3	11	34.4	5	2	B44817
4	11	34.4	5	2	D44817
5	10	31.2	7	2	E30608
6	9	28.1	3	2	T13892
7	9	28.1	5	2	E42364
8	9	28.1	5	2	E60274
9	9	28.1	5	2	A37114
10	9	28.1	5	2	H44817
11	9	28.1	5	2	F44817
12	9	28.1	5	2	S69237
13	9	28.1	6	2	A03986
14	9	28.1	6	2	S14159
15	9	28.1	6	2	A43766
16	9	28.1	6	2	I37263
17	9	28.1	6	2	I65546
18	9	28.1	7	2	S25266
19	9	28.1	7	2	A25269
20	9	28.1	7	2	A30812
21	9	28.1	7	2	PM0649
22	9	28.1	7	2	PM0646
23	9	28.1	7	2	PM0644
24	8	25.0	3	3	PT0622
25	8	25.0	4	3	PT0696
26	8	25.0	4	3	PT0645
27	8	25.0	4	3	PT0712
28	8	25.0	4	3	PT0698
29	8	25.0	4	2	PT0551

#### ALIGNMENTS

30	8	25.0	4	2	PT0697	T-cell receptor be
31	8	25.0	5	2	A05321	glycogen phosphory
32	8	25.0	5	2	I40698	biotin B - Citropa
33	8	25.0	5	2	I39964	ribosomal protein
34	8	25.0	5	2	I39966	ribosomal protein
35	8	25.0	5	2	I39965	ribosomal protein
36	8	25.0	5	2	T10954	hypothetical prote
37	8	25.0	5	2	I50385	myosin light chain
38	8	25.0	5	2	PT0308	Ig heavy chain CRD
39	8	25.0	5	2	PT0596	T-cell receptor be
40	8	25.0	5	2	PT0610	T-cell receptor be
41	8	25.0	5	2	PT0597	T-cell receptor be
42	8	25.0	5	2	PT0600	T-cell receptor be
43	8	25.0	5	2	PT0729	T-cell receptor be
44	8	25.0	5	2	PT0624	T-cell receptor be
45	8	25.0	5	2	PT0686	T-cell receptor be

RESULT 1  
A40135  
branched-chain-amino-acid transaminase (EC 2.6.1.42), mitochondrial - rat (fragment)  
N:Alternate names: branched-chain-amino-acid aminotransferase, mitochondrial  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 13-May-1992 #sequence\_revision 13-May-1992 #text\_change 30-Sep-1993  
C:Accession: A40135  
R:Hutson, S.W.; Wallin, R.; Hall, T.R.  
Submitted to the Protein Sequence Database, March 1992  
A:Reference number: A40135  
A:Accession: A40135  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-4 <HUT>  
C:Keywords: aminotransferase; mitochondrion

Query Match 34.4% Score 11; DB 2; Length 4;  
Best Local Similarity 66.7% Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 ISS 7  
DB 1 VSS 3  
RESULT 2  
PT0644  
T-cell receptor beta chain V-D-J region (111-16) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
C:Accession: PT0644  
R:Peeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A:Title: Functional sequences of fetal T cell receptor beta chains have few N regions  
A:Reference number: PT0509; MUID:9177601  
A:Accession: PT0644  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-5 <PEE>  
A:Experimental source: newborn thymus, strain BALB/c  
C:Keywords: T-cell receptor

Query Match 34.4% Score 11; DB 2; Length 5;  
Best Local Similarity 100.0% Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FT 2  
DB 4 FT 5

RESULT 3  
B44817  
34.5K structural protein - Leuconostoc oenos phage p2L11-15 (fragment)  
C:Species: Leuconostoc oenos phage p2L11-15  
C:Date: 31-Mar-1993 #sequence\_revision 22-May-1998 #text\_change 22-May-1998  
C:Accession: B44817  
R:Arndt, E.K.; Lonvaud, A.; Hammes, W.P.  
J. Gen. Microbiol. 137, 2135-2139, 1991  
A:Title: Lysogeny in Leuconostoc oenos.  
A:Reference number: A44817; MUID:92085033  
A:Accession: B44817  
A:Molecule type: protein  
A:Residues: 1-5 <ABE>  
A:Note: sequence extracted from NCBI backbone (NCBIP:70342)

Query Match 34.4%; Score 11; DB 2; Length 5;  
Best Local Similarity 60.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 3 LTRSS 7  
Db 1 LATSS 5

RESULT 4  
D44817  
35K structural protein - Leuconostoc oenos phage pA5-12 (fragment)  
C:Species: Leuconostoc oenos phage pA5-12  
C:Date: 31-Mar-1993 #sequence\_revision 22-May-1998 #text\_change 22-May-1998  
C:Accession: D44817  
R:Arndt, E.K.; Lonvaud, A.; Hammes, W.P.  
J. Gen. Microbiol. 137, 2135-2139, 1991  
A:Title: Lysogeny in Leuconostoc oenos.  
A:Reference number: A44817; MUID:92085033  
A:Accession: D44817  
A:Molecule type: protein  
A:Residues: 1-5 <ABE>  
A:Note: sequence extracted from NCBI backbone (NCBIP:70340)

Query Match 34.4%; Score 11; DB 2; Length 5;  
Best Local Similarity 60.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 3 LTRSS 7  
Db 1 LATSS 5

RESULT 5  
E30608  
19 Kappa chain V-III region (dag) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 29-Jun-1989 #sequence\_revision 29-Jun-1989 #text\_change 16-Aug-1996  
C:Accession: E30608  
R:Goni, F.R.; Chen, P.P.; McGinnis, D.; Arjonilla, M.L.; Fernandez, J.; Carson, D.; Sold  
J. Immunol. 142, 3158-3163, 1989  
A:Title: Structural and idiotypic characterization of the L chains of human Igm autocalt  
A:Reference number: A30601; MUID:89215279  
A:Accession: E30608  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-7 <GON>  
C:Keywords: heterotrimer; immunoglobulin

Query Match 31.2%; Score 10; DB 2; Length 7;  
Best Local Similarity 75.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 3 LTRSS 6

Db 4 LTRSS 7

RESULT 6  
T13892  
cytochrome-c oxidase (EC 1.9.3.1) chain I (imported) - river lamprey mitochondrion (f  
C:Species: mitochondrion lampetra fluviatilis (river lamprey)  
C:Date: 15-Jun-2001 #sequence\_revision 15-Jun-2001 #text\_change 15-Jun-2001  
C:Accession: T13892  
R:Delarbre, C.; Barriel, V.; Tiller, S.; Janvier, P.; Gachelin, G.  
Mol. Biol. Evol. 14, 807-815, 1997  
A:Title: The main features of the granite mitochondrial DNA between the NDI and the  
A:Reference number: Z17775; MUID:97398704  
A:Accession: T13892  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3 <DEL>  
A:Cross-references: EMBL:Y09528; NID:92340016; PIDN:CAA70721.1; PID:94379123  
C:Genetics:  
A:Genome: mitochondrion  
A:Note: COI  
C:Keywords: mitochondrion; oxidoreductase

Query Match 28.1%; Score 9; DB 3; Length 3;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 TL 3  
Db 2 TL 3

RESULT 7  
E42364  
flagellar protein flir - Salmonella typhimurium (fragment)  
C:Species: Salmonella typhimurium  
C:Date: 24-Jul-1992 #sequence\_revision 24-Jul-1992 #text\_change 30-Sep-1993  
C:Accession: E42364  
R:Voelger, A.P.; Homma, M.; Irikura, V.M.; Macnab, R.M.  
J. Bacteriol. 173, 3564-3572, 1991  
A:Title: Salmonella typhimurium mutants defective in flagellar filament regrowth and  
A:Reference number: A42364; MUID:91258342  
A:Accession: E42364  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-5 <YOG>  
A:Cross-references: GB:M62408

Query Match 28.1%; Score 9; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 TL 3  
Db 3 TL 4

RESULT 8  
E60274  
major protein antigen MP763 - Mycobacterium tuberculosis (fragment)  
C:Species: Mycobacterium tuberculosis  
C:Date: 11-Dec-1992 #sequence\_revision 11-Dec-1992 #text\_change 30-Sep-1993  
C:Accession: E60274  
R:Nagai, S.; Wiker, H.G.; Harboe, M.; Klnomoto, M.  
Infect. Immun. 59, 372-382, 1991  
A:Title: Isolation and partial characterization of major protein antigens in the cult  
A:Reference number: A60274; MUID:91099989  
A:Accession: E60274  
A>Status: preliminary  
A:Molecule type: protein



A:Residues: 1-5 <NMG>

Query Match 28.1%; Score 9; DB 2; Length 5;  
Best Local Similarity 25.0%; Pred. No. 2.8e+05;  
Matches 1; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTLT 4  
: |  
Db 2 YPIT 5

RESULT 9

A37114 hypoxanthine phosphoribosyltransferase (EC 2.4.2.8) - fluke (Schistosoma mansoni) (fragm

C:Species: Schistosoma mansoni  
C:Date: 28-Mar-1991 #sequence\_revision 28-Mar-1991 #text\_change 23-Jun-1993

C:Accession: A37114

R:Yuan, L.; Craig, S.P.; McKerrow, J.H.; Wang, C.C.

J. Biol. Chem. 265, 13528-13532, 1990

A:Title: The hypoxanthine-guanine phosphoribosyltransferase of Schistosoma mansoni. Furth

A:Reference number: A37114; MUID:90337955

A:Accession: A37114

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-5 <YUA>

C:Keywords: glycosyltransferase; pentosyltransferase

Query Match 28.1%; Score 9; DB 2; Length 5;  
Best Local Similarity 66.7%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 5 ISS 7  
: |  
Db 1 MSS 3

RESULT 10

H44817 34.5K structural protein - Leuconostoc oenos phage P32 (fragment)

C:Species: Leuconostoc oenos phage P32  
C:Date: 31-Mar-1993 #sequence\_revision 22-May-1998 #text\_change 22-May-1998

C:Accession: H44817

R:Arndt, E.K.; Lonvaud, A.; Hammes, W.P.

J. Gen. Microbiol. 137, 2135-2139, 1991

A:Title: Lysozyme in Leuconostoc oenos

A:Reference number: A44817; MUID:92085033

A:Molecule type: protein

A:Residues: 1-5 <ARE>

A:Note: sequence extracted from NCBI backbone (NCBIP:70332)

Query Match 28.1%; Score 9; DB 2; Length 5;  
Best Local Similarity 40.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 LTISS 7  
: |  
Db 1 MATSS 5

RESULT 11

F44817 34.5K structural protein - Leuconostoc oenos phage P34 (fragment)

C:Species: Leuconostoc oenos phage P34  
C:Date: 31-Mar-1993 #sequence\_revision 22-May-1998 #text\_change 22-May-1998

C:Accession: F44817

R:Arndt, E.K.; Lonvaud, A.; Hammes, W.P.

J. Gen. Microbiol. 137, 2135-2139, 1991

A:Title: Lysozyme in Leuconostoc oenos

A:Reference number: A44817; MUID:92085033

A:Accession: F44817  
A:Molecule type: protein  
A:Residues: 1-5 <ARE>  
A:Note: sequence extracted from NCBI backbone (NCBIP:70335)

Query Match 28.1%; Score 9; DB 2; Length 5;  
Best Local Similarity 40.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 LTISS 7  
: |  
Db 1 MATSS 5

RESULT 12

S69237 surface protein tetrabrachion heavy chain - Staphylothermus marinus (fragment)

C:Species: Staphylothermus marinus

C:Date: 04-Dec-1997 #sequence\_revision 04-Dec-1997 #text\_change 17-Mar-1999

C:Accession: S69237

R:Petters, U.; Nilsen, M.; Kuehlmoorgen, B.; Goldik, R.; Lupas, A.; Kellermann, J.; Eng

J. Mol. Biol. 245, 385-401, 1995

A:Title: Tetrabrachion: a filamentous archaeobacterial surface protein assembly of unu

A:Reference number: S69237; MUID:95139068

A:Accession: S69237

A:Molecule type: protein

A:Residues: 1-5 <PER>

A:Experimental source: strain FL, DSM 3639

C:Keywords: cell wall; glycoprotein; heat-stable protein

Query Match 28.1%; Score 9; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TL 3  
: |  
Db 2 TL 3

RESULT 13

A60986 N-formyl oligopeptide - Escherichia coli (fragment)

C:Species: Escherichia coli

C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 31-Dec-1993

C:Accession: A60986

R:Broom, M.F.; Mellor, D.M.; Chadwick, V.S.

Experientia 45, 1097-1099, 1989

A:Title: Purification and amino acid sequencing of naturally occurring N-formyl-methi

A:Reference number: A60986; MUID:90092408

A:Accession: A60986

A:Molecule type: protein

A:Residues: 1-6 <BRO>

C:Comment: This hexapeptide was the longest of several N-formyl oligopeptides reporte

F:1/modified site: N-formylmethionine #status experimental

Query Match 28.1%; Score 9; DB 2; Length 6;  
Best Local Similarity 66.7%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTL 3  
: |  
Db 3 FTL 5

RESULT 14

S14159 parasporal crystal protein CRYIC - Bacillus thuringiensis (fragment)

N:Alternate names: delta-endotoxin

C:Species: Bacillus thuringiensis

C:Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 18-Jul-1997

C:Accession: S14159  
R:Convents, D.; Cherlet, M.; van Damme, J.; Lasters, I.; Lauwereys, M.  
Eur. J. Biochem. 195, 631-635, 1991  
A:Title: Two structural domains as a general fold of the toxic fragment of the Bacillus  
A:Reference number: S14087; MUID:91153300  
A:Accession: S14159  
A:Molecule type: protein  
A:Residues: 1-6 <CON>

Query Match 28.1%; Score 9; DB 2; Length 6;  
Best local similarity 66.7%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 ISS 7  
11  
Db 1 IST 3

## RESULT 15

A43766

28K ubiquitin-immunoreactive protein - inxy cap (Coprinus cinereus) (fragment)

C:Species: Coprinus cinereus

C:Date: 01-Dec-1992 #sequence\_revision 01-Dec-1992 #text\_change 30-Sep-1993

C:Accession: A43766

R:Kanda, T.; Inoue, M.; Akiyama, M.

Biochimie 72, 355-359, 1990

A:Title: Purification and characterization of an ubiquitin-immuno-reactive protein local

A:Reference number: A43766; MUID:91002724

A:Accession: A43766

A&gt;Status: preliminary

A:Molecule type: protein

A:Residues: 1-6 &lt;KAN&gt;

Query Match 28.1%; Score 9; DB 2; Length 6;  
Best local similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TL 3  
11  
Db 5 TL 6

Search completed: July 15, 2002, 13:27:07  
Job time: 437 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:26:20 : Search time 24.49 Seconds  
(without alignments)  
11.067 Million cell updates/sec

Title: US-09-712-819A-1

Perfect score: 32

Sequence: 1 FITTSS 7

Scoring table: BIOSM62

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 84

Minimum DB seq length: 0

Maximum DB seq length: 7

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing First 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Length	DB	ID	Description
1	12	37.5	7	1	GPR_MOUSE
2	11	34.4	5	1	RE31_LITRU
3	11	34.4	5	1	RE32_LITRU
4	10	31.2	5	1	E104_LITRU
5	9	28.1	7	1	CCFL_ENTFA
6	9	28.1	7	1	C1A_ENTFA
7	8	25.0	5	1	AL14_CARMA
8	8	25.0	5	1	BI0B_CITFR
9	8	25.0	5	1	PSK_DAVCA
10	8	25.0	6	1	LOK1_LOCOMI
11	8	25.0	7	1	AL12_CARMA
12	8	25.0	7	1	AL13_CARMA
13	8	25.0	7	1	AL14_CARMA
14	8	25.0	7	1	AL15_CARMA
15	8	25.0	7	1	ALT7_CITPO
16	8	25.0	7	1	UC24_MAIZE
17	7	21.9	4	1	BOS1_HUMAN
18	7	21.9	4	1	RM01_YEAST
19	7	21.9	6	1	B10A_CITFR
20	7	21.9	6	1	UN06_CLOPA
21	6	18.8	4	1	ACH3_ACHFU
22	6	18.8	4	1	FAR3_HTRME
23	6	18.8	4	1	FAR4_HTRME
24	6	18.8	4	1	FEKA_ANTPEI
25	6	18.8	4	1	FLRE_HTRME
26	6	18.8	4	1	FLRE_HTRME
27	6	18.8	4	1	FLRE_HTRME
28	6	18.8	4	1	FLRE_HTRME
29	6	18.8	4	1	CCP1_OCTMI
30	6	18.8	5	1	E103_LITRU
31	6	18.8	5	1	FARP_ARTTR
32	6	18.8	5	1	PAP2_PARMA
33	6	18.8	5	1	RE11_LITRU

34	6	18.8	5	1	RE21_LITRU	P82071	litorea rub
35	6	18.8	5	1	SUG5_ACHDO	P19991	acheta dome
36	6	18.8	5	1	TPIS_CARMA	P54714	caeta fam1
37	6	18.8	5	1	UC22_MAIZE	P80628	zea mays (m
38	6	18.8	6	1	C1P1_MYTED	P13736	mytilus edu
39	6	18.8	6	1	C1P2_MYTED	P13737	mytilus edu
40	6	18.8	6	1	E101_LITRU	P82096	litorea rub
41	6	18.8	6	1	FARP_MONEX	P41966	moniezia ex
42	6	18.8	7	1	FAR1_ASCSU	P31889	ascaris suu
43	6	18.8	7	1	FAR1_HELTI	P41871	helisoma tr
44	6	18.8	7	1	FAR1_PROCL	P38459	procambarus
45	6	18.8	7	1	FAR2_ASCSU	P31890	ascaris suu

## ALIGNMENTS

RESULT 1

GPR\_MOUSE STANDARD; PRT: 7 AA.

AC P99025;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE GTP cyclohydrolase I feedback regulatory protein (P35) (Fragment).

GN GCHFR OR GPRP.

OS Mus musculus (mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE.

RC TISSUE=Liver;

RA Sanchez J.-C., Rouge V., Frutiger S., Hughes G., Yan J.X., Hoogland C., Appel R.D., Binz P.-A., Hochstrasser D.F., Cowthorne M.;

RA Submitted (Aug-1998) to the SWISS-PROT data bank.

CC -! FUNCTION: MEDIATES TETRAHYDROBIPTERIN INHIBITION OF GTP CYCLOHYDROLASE I. THIS INHIBITION IS REVERSED BY L-PHENYLANILINE (BY SIMILARITY).

CC -! SUBUNIT: HOMODIMER (BY SIMILARITY).

CC SWISS-2DPAGE: P99025; MOUSE.

DR INIT\_MET

FT NON\_TER

SQ SEQUENCE 7 AA; 806 MW; 71B5B057273B4700 CRC64;

Query Match 37.5%; Score 12; DB 1; Length 7;

Best Local Similarity 60.0%; Pred. No. 1e+05;

Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 LTSS 7

DB 3 LIST 7

RESULT 2

RE31\_LITRU STANDARD; PRT: 5 AA.

AC P82072;

DT 01-MAR-2002 (Rel. 41, Created)

DT 01-MAR-2002 (Rel. 41, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Rubellididn 3.1.

OS Litorea rubella (desert tree frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae; Litorea.

OX NCBI\_TaxID=104895;

RN [1]

RC TISSUE=Skin secretion;

RC Steinboerner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,

RA Tyler M.J., Wallace J.C.;  
 RT "The structure of new peptides from the Australian red tree frog  
 RT 'Litoria rubella', the skin peptide profile as a probe for the study  
 RT of evolutionary trends of amphibians.";  
 RL Aust. J. Chem. 49:955-963(1996).  
 CC -1- FUNCTION: SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC  
 CC ACTIVITY.  
 CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
 CC -1- MASS SPECTROMETRY: MW=655; METHOD=FAB.  
 KW Amphibian skin: Amelation.  
 FT MOD.RES 5  
 SQ SEQUENCE 5 AA: 656 MW: 71A9C9CB10300000 CRC64:

Query Match 34.48; Score 11; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FT 2  
 II  
 DB 4 FT 5

RESULT 3  
 RE32\_LITRU STANDARD; PRT; 5 AA.  
 ID RE32\_LITRU  
 AC P82073;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Rubellidin 3.2.  
 OS Litoria rubella (Desert tree frog).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
 CC Litoria.  
 OX NCBI\_TaxID=104895;  
 RN [1]  
 RP SEQUENCE.  
 RC Tissue-skin secretion;  
 RA Wabnitz P.A., Bowle J.H., Tyler M.J., Wallace J.C.;  
 RT "Peptides from the skin glands of the Australian buzzing tree frog  
 RT Litoria electrica. Comparison with the skin peptides from Litoria  
 RT rubella.";  
 RL Aust. J. Chem. 52:0-0(1999).  
 CC -1- FUNCTION: SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC  
 CC ACTIVITY.  
 CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
 KW Amphibian skin.  
 SQ SEQUENCE 5 AA: 570 MW: 71A9C9C862A00000 CRC64:

Query Match 34.48; Score 11; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FT 2  
 II  
 DB 4 FT 5

RESULT 4  
 E104\_LITRU STANDARD; PRT; 5 AA.  
 ID E104\_LITRU  
 AC P82100;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Electrin 4.  
 OS Litoria rubella (Desert tree frog).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
 CC Litoria.  
 OX NCBI\_TaxID=104895;

RN [1]  
 RP SEQUENCE.  
 RC Tissue-skin secretion;  
 RA Wabnitz P.A., Bowle J.H., Tyler M.J., Wallace J.C.;  
 RT "Peptides from the skin glands of the Australian buzzing tree frog  
 RT Litoria electrica. Comparison with the skin peptides from Litoria  
 RT rubella.";  
 RL Aust. J. Chem. 52:0-0(1999).  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC Amphibian skin: Amelation.  
 FT MOD.RES 5  
 SQ SEQUENCE 5 AA: 616 MW: 61F2D1A059A00000 CRC64:

Query Match 31.28; Score 10; DB 1; Length 5;  
 Best Local Similarity 33.3%; Pred. No. 1e+05;  
 Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 LTI 5  
 II  
 DB 2 ITV 4

RESULT 5  
 CCE1\_ENTFA STANDARD; PRT; 7 AA.  
 ID CCE1\_ENTFA  
 AC P20104;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-FEB-1991 (Rel. 17, Last annotation update)  
 DE Sex pheromone CCF10.  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;  
 CC Enterococcus.  
 OX NCBI\_TaxID=1351;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=89008313; PubMed=3139658;  
 RA Mori M., Sakagami Y., Ishii Y., Isogai A., Kitada C., Fujino M.,  
 RA Abele J.C., Dunn G.M., Suzuki A.;  
 RT "Structure of CCF10, a peptide sex pheromone which induces  
 RT conjugative transfer of the Streptococcus faecalis tetracycline  
 RT resistance plasmid, pCF10.";  
 RL J. Biol. Chem. 263:14574-14578(1988).  
 CC -1- FUNCTION: CCF10 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE  
 CC HEMOLYSIN PLASMID PCF10.  
 KW PIR; A30812; A30812.  
 SQ SEQUENCE 7 AA: 790 MW: 72C9D2C731B2C740 CRC64:

Query Match 28.18; Score 9; DB 1; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TL 3  
 II  
 DB 3 TL 4

RESULT 6  
 C1A\_ENTFA STANDARD; PRT; 7 AA.  
 ID C1A\_ENTFA  
 AC P11932;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 01-FEB-1991 (Rel. 17, Last annotation update)  
 DE Sex pheromone CAM373 (Clumping-inducing agent) (CIA).  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;  
 CC Enterococcus.  
 OX NCBI\_TaxID=1351;  
 RN [1]



SO SEQUENCE 5 AA: 687 MW: 76C1BB50AB300000 CRC64:

Query Match 25.0%; Score 8; DB 1; Length 5;  
Best Local Similarity 50.0%; Pred. No. 1e+05;  
Matches 1: Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FT 2  
1  
3 YL 4

RESULT 10

LOC1 LOCMT STANDARD; PRT: 6 AA.  
ID LOC1 LOCMT  
AC P41491;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Locustakinin 1.  
OS Locusta migratoria (Migratory locust).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;  
OC Acridoidea; Acridoidea; Acrididae; Locusta.  
OX NCBI\_Taxid=7004;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Corpora cardiaca;  
RA MEDLINE=92262851; PubMed=1585017;  
RA Schoofs L., Holman G.M., Preost P., van Damme J., Hayes T.K.,  
de Loof A.;  
RT "Locustakinin, a novel myotropic peptide from locusta migratoria,  
RT Isolation, primary structure and synthesis.";  
RL Regul. Pept. 37:49-57(1992).  
CC -1- FUNCTION: MYOTROPIC PEPTIDE. MAY BE IMPORTANT IN THE STIMULATION  
CC OF ION TRANSPORT AND INHIBITION OF DIURETIC ACTIVITY IN MALPIGHIAN  
CC TUBULES.  
CC PIR: A61068.  
DR Neuropeptide; Amidation.  
KW MOD\_RES 6  
FT SEQUENCE 6 AA: 654 MW: 686365A5B9CDB000 CRC64:

Query Match 25.0%; Score 8; DB 1; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1e+05;  
Matches 2: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 SS 7  
1  
3 SS 4

RESULT 11

AL12 CARMA STANDARD; PRT: 7 AA.  
ID AL12 CARMA  
AC P81805;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Carcinustatin 2.  
OS Carcinus maenas (Common shore crab) (Green crab).  
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
OX NCBI\_Taxid=6759;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
RA MEDLINE=98121193; PubMed=9461295;  
RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
Thorpe A.;  
RT Isolation and identification of multiple neuropeptides of the  
RT allatostylin superfamily in the shore crab Carcinus maenas.";

RL Eur. J. Biochem. 250:727-734(1997)  
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
KW Neuropeptide; Amidation; Multigene family.  
FT MOD\_RES 7  
SQ SEQUENCE 7 AA: 770 MW: 672879CDB5DB70 CRC64;

Query Match 25.0%; Score 8; DB 1; Length 7;  
Best Local Similarity 66.7%; Pred. No. 1e+05;  
Matches 2: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTL 3  
1  
5 FGL 7

RESULT 12

AL13 CARMA STANDARD; PRT: 7 AA.  
ID AL13 CARMA  
AC P81806;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Carcinustatin 3.  
OS Carcinus maenas (Common shore crab) (Green crab).  
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
OX NCBI\_Taxid=6759;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
RA MEDLINE=98121193; PubMed=9461295;  
RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
Thorpe A.;  
RT "Isolation and identification of multiple neuropeptides of the  
RT allatostylin superfamily in the shore crab Carcinus maenas.";  
RL Eur. J. Biochem. 250:727-734(1997).  
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
KW Neuropeptide; Multigene family.  
SQ SEQUENCE 7 AA: 796 MW: 672879CDB476B70 CRC64:

Query Match 25.0%; Score 8; DB 1; Length 7;  
Best Local Similarity 66.7%; Pred. No. 1e+05;  
Matches 2: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTL 3  
1  
5 FGL 7

RESULT 13

AL14 CARMA STANDARD; PRT: 7 AA.  
ID AL14 CARMA  
AC P81807;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Carcinustatin 4.  
OS Carcinus maenas (Common shore crab) (Green crab).  
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
OX NCBI\_Taxid=6759;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
RA MEDLINE=98121193; PubMed=9461295;  
RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
Thorpe A.;

\* Isolation and identification of multiple neuropeptides of the  
 RT allatostatin superfamily in the shore crab *Carcinus maenas*.";  
 RL Eur. J. Biochem. 250:727-734(1997).  
 CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 KW Neuropeptide; Multigene family.  
 SQ SEQUENCE 7 AA: 782 MW; 672879CDBCBA76AC0 CRC64;

Query Match 25.0%; Score 8; DB 1; Length 7;  
 Best Local Similarity 66.7%; Pred. No. 1e+05;  
 Matches 2: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTL 3  
 | |  
 DB 5 FGL 7

RESULT 14  
 ID ALL5\_CARMA STANDARD; PRT; 7 AA.  
 AC P81808;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Carcinustatin 5.  
 OS *Carcinus maenas* (Common shore crab) (Green crab).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 CC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
 OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
 OX NCBI\_TaxID=6759;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
 RX MEDLINE=98121193; PubMed=9461295;  
 RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
 RA Thorpe A.;  
 RT "Isolation and identification of multiple neuropeptides of the  
 RT allatostatin superfamily in the shore crab *Carcinus maenas*.";  
 RL Eur. J. Biochem. 250:727-734(1997).  
 CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 KW Neuropeptide; Amidation; Multigene family.  
 FT MOD\_RES 7  
 SQ SEQUENCE 7 AA: 781 MW; 672879CDBCBA76420 CRC64;

Query Match 25.0%; Score 8; DB 1; Length 7;  
 Best Local Similarity 66.7%; Pred. No. 1e+05;  
 Matches 2: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTL 3  
 | |  
 DB 5 FGL 7

RESULT 15  
 ID ALL7\_CYPDPO STANDARD; PRT; 7 AA.  
 AC P82158;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Cydiastatin 7.  
 OS *Cydia pomonella* (Codling moth).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Tortricoidae; Tortricidae; Olethreutinae; Cydia.  
 OX NCBI\_TaxID=82600;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Larva;  
 RX MEDLINE=98054539; PubMed=9392829;

RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,  
 RA Davey M., East P.D., Thorpe A.;  
 RT "Lepidopteran peptides of the allatostatin superfamily.";  
 RL Peptides 18:1301-1309(1997).  
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 KW Neuropeptide; Amidation.  
 FT MOD\_RES 7  
 SQ SEQUENCE 7 AA: 873 MW; 672879CABB569350 CRC64;

Query Match 25.0%; Score 8; DB 1; Length 7;  
 Best Local Similarity 66.7%; Pred. No. 1e+05;  
 Matches 2: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTL 3  
 | |  
 DB 5 FGL 7

Search completed: July 15, 2002, 13:38:10  
 Job time: 710 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 15, 2002, 13:25:30 ; Search time 86.73 Seconds

(without alignments)  
13.962 Million cell updates/sec

Title: US-09-712-819A-1

Perfect score: 32

Sequence: 1 PRLTIS 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 65

Minimum DB seq length: 0

Maximum DB seq length: 7

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_PROTOZOA:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_EUKARYOT:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	50.0	7	2 P70804	P70804 azotobacter
2	12	37.5	7	2 O34028	O34028 spinaciola
3	11	34.4	5	13 P82072	P82072 litoria rub
4	11	34.4	5	13 P82073	P82073 litoria rub
5	11	34.4	7	2 O54248	O54248 streptococ
6	10	31.2	5	13 P82100	P82100 litoria rub
7	9	28.1	7	2 O07354	O07354 streptococ
8	9	28.1	7	10 P82445	P82445 streptococ
9	8	25.0	6	10 P82181	P82181 spinaciola
10	8	25.0	6	10 P82182	P82182 spinaciola
11	8	25.0	7	2 P72081	P72081 nocardia
12	7	21.9	7	2 O47029	O47029 enterobacte
13	7	21.9	7	12 O65578	O65578 bovine herp
14	6	18.8	5	13 P82070	P82070 litoria rub
15	6	18.8	5	13 P82071	P82071 litoria rub
16	6	18.8	5	13 P82099	P82099 litoria rub

17	6	18.8	6	10 P82541	P82541 spinaciola
18	6	18.8	6	13 P82096	P82096 litoria rub
19	6	18.8	7	2 O50556	O50556 actinobacill
20	6	18.8	7	4 O15903	O15903 homo sapien
21	6	18.8	7	8 O95945	O95945 saccharomyc
22	6	18.8	7	8 O98866	O98866 spinaciola
23	6	18.8	7	10 O9C5B3	O9C5B3 aradidopsis
24	6	18.8	7	10 O9C5B3	O9C5B3 aradidopsis
25	6	18.8	7	13 O42564	O42564 ligu rubrip
26	6	18.8	7	13 P82065	P82065 litoria rub
27	5	15.6	5	2 P83073	P83073 bacillus ce
28	5	15.6	7	2 O47505	O47505 escherichia
29	5	15.6	7	10 P93233	P93233 lycopersico
30	5	15.6	7	11 O63668	O63668 rattus norv
31	5	15.6	7	12 O67113	O67113 influenza a
32	5	15.6	7	13 P82101	P82101 litoria rub
33	5	15.6	7	15 O07624	O07624 rous sarcom
34	4	12.5	4	11 O08833	O08833 rattus norv
35	4	12.5	7	4 O15897	O15897 homo sapien
36	4	12.5	7	8 P92421	P92421 psathyrosta
37	4	12.5	7	8 P92385	P92385 hordeum mar
38	4	12.5	7	8 P92372	P92372 haynaldia v
39	4	12.5	7	8 P92403	P92403 lophopyrum
40	4	12.5	7	8 P92425	P92425 pseudoroegn
41	4	12.5	7	8 P92387	P92387 henardia p
42	4	12.5	7	8 P92427	P92427 peridictyon
43	4	12.5	7	8 P92390	P92390 heteranthel
44	4	12.5	7	8 P92226	P92226 crithopsis
45	4	12.5	7	8 P92214	P92214 amblyopyrum

## ALIGNMENTS

RESULT 1  
ID P70804 PRELIMINARY; PRT; 7 AA.  
AC P70804:  
DT 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ALGT PROTEIN (FRAGMENT).  
GN ALGT.  
OS Azotobacter vinelandii.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Azotobacter.  
OX NCBI\_TaxID=354;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=E;  
RX MEDLINE=96427318; PubMed=8830682;  
RA Rehm B.H.A., Ertesvag H., Valla S.;  
RT "A new Azotobacter vinelandii mannuram C-5-epimerase gene (algG) is  
RT to that in Pseudomonas aeruginosa.";  
RT J. Bacteriol. 178:5884-5889 (1996).  
DR EMBL; X87973; CAA61230.1; -;  
FT NON\_TER  
SQ SEQUENCE 7 AA: 684 MW: 7185A5A5A2D1AED0 CRC64;

Query Match 50.0%; Score 16; DB 2; Length 7;  
Best Local Similarity 75.0%; Pred. No. 5.6e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 TISS 7  
DB 2 TVSS 5  
RESULT 2  
ID O34028 PRELIMINARY; PRT; 7 AA.

AC 034028;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 13, Last annotation update)  
DE CATECHOL-2,3-DIOXYGENASE (FRAGMENT).  
GN PHE.  
OS Sphingomonas chungbukensis.  
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;  
CC Sphingomonas.  
NCBI\_TaxID=56193;  
RN [1]  
RP SPOUNCE FROM N.A.  
RC STRAIN-DJ77;  
RA Kim Y.-C.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; 088298; AAB6511.1; -.  
KW Dioxxygenase.  
FT NON\_TER  
SQ SEQUENCE 7 AA; 868 MW; 71A452D1A699D460 CRC64;

Query Match 37.5%; Score 12; DB 2; Length 7;  
Best Local Similarity 20.0%; Pred. No. 5.6e+05;  
Matches 1; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 3 LTSS 7  
DB 3 MYNT 7

RESULT 3  
P82072 PRELIMINARY; PRT; 5 AA.  
AC P82072;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE ROBELLDIN 3.1.  
OS Litoria rubella (Desert tree frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litoria.  
NCBI\_TaxID=104895;  
RN [1]  
RP SEQUENCE, AND MASS SPECTROMETRY.  
RC TISSUE-SKIN SECRETION.  
RA Streinborner S.T., Mahlitz P.A., Maugh R.J., Bowie J.H., Gao C.,  
RA Tyler M.J., Wallace J.C.;  
RT 'The structure of new peptides from the Australian red tree frog  
RT 'Litoria rubella', the skin peptide profile as a probe for the study  
RT of evolutionary trends of amphibians.';  
RL Aust. J. Chem. 49:955-963(1996).  
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR  
CC ANTIBIOTIC ACTIVITY.  
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
CC -1- MASS SPECTROMETRY: MW=655; METHOD=FAH.  
KM Amphibian skin; Amelation.  
FT MOD\_RES  
SQ SEQUENCE 5 AA; 656 MW; 71A9C9CB10300000 CRC64;

Query Match 34.4%; Score 11; DB 13; Length 5;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FT 2  
DB 4 FT 5

RESULT 4  
P82073 PRELIMINARY; PRT; 5 AA.

AC P82073;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE RUDELIDIN 3.2.  
OS Litoria rubella (Desert tree frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litoria.  
NCBI\_TaxID=104895;  
RN [1]  
RP SEQUENCE.  
RC TISSUE-SKIN SECRETION.  
RA Maunitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;  
RT 'Peptides from the skin glands of the Australian buzzing tree frog  
RT Litoria electrica. Comparison with the skin peptides from Litoria  
RT rubella.';  
RL Aust. J. Chem. 52:0-0(1999).  
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR  
CC ANTIBIOTIC ACTIVITY.  
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
KM Amphibian skin.  
FT NON\_TER  
SQ SEQUENCE 5 AA; 570 MW; 71A9C9CB62A00000 CRC64;

Query Match 34.4%; Score 11; DB 13; Length 5;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FT 2  
DB 4 FT 5

RESULT 5  
O54248 PRELIMINARY; PRT; 7 AA.  
AC O54248;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE RPLO PROTEIN (FRAGMENT).  
GN RPLO.  
OS Streptomyces griseus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
CC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
NCBI\_TaxID=1911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-N2-3-11;  
RX MEDLINE=20011291; PubMed=10542330;  
RA Poehling S., Piepersberg W., Wehmeler U.F.;  
RT 'Analysis and interaction of the sec Y gene from Streptomyces griseus  
RT N2-3-11 and interaction of the Sec Y protein with the SecA protein.';  
RL Blochm. Biophys. Acta 1447:298-302(1999).  
DR EMBL; X95915; CAA65160.1; -.  
FT NON\_TER  
SQ SEQUENCE 7 AA; 760 MW; 72C72B01B2D1B2A0 CRC64;

Query Match 34.4%; Score 11; DB 2; Length 7;  
Best Local Similarity 66.7%; Pred. No. 5.6e+05;  
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 TLT 4  
DB 2 TLT 4

RESULT 6  
P82100 PRELIMINARY; PRT; 5 AA.  
AC P82100;

DT 01-MAY-2000 (TREMBLrel. 13. Created)  
DT 01-MAY-2000 (TREMBLrel. 13. Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13. Last annotation update)  
DE ELECTRIN 4  
OS Litorea rubella (Desert tree frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litorea.  
OX NCBI\_TaxID=104895;  
RN [1]  
RP  
RP TISSUE=SKIN SECRETION;  
RA Mabinitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;  
RT "Peptides from the skin glands of the Australian buzzing tree frog  
RT *Litoria electrica*. Comparison with the skin peptides from *Litoria*  
RT *rubella*.";  
RT Aust. J. Chem. 52:0-0(1999).  
RL  
RL Amphibian skin; Amidation.  
RT MOD\_RES 5  
SQ SEQUENCE 5 AA; 616 MW; 612PDAD059A00000 CRC64;

	Query Match	31.2%	Score 10:	DB 13:	Length 5:
	Best Local Similarity	33.3%	Pred. No.	5.6e+05:	
	Matches	1;	Conservative	2;	Mismatches 0; Indels 0; Gaps 0;
OY	3 LRT 5				
	: 1:				
Dd	2 RTV 4				

RESULT	7			
ID	007354	PRELIMINARY;	PRT:	7 AA.
AC	007354;			
DT	01-JUL-1997 (TREMBLrel_04, Created)			
DT	01-JUL-1997 (TREMBLrel_04, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel_19, Last annotation update)			
DE	NIFK (FRAGMENT).			
GN	NIFK.			
OS	Synechococcus sp. (strain PCC 8801 / RF-1) (Cyanothecae PCC 8801).			
OC	Bacteria; Cyanobacteria; Chroococcales; Cyanothecae.			
OX	NCBI_TaxID=41431;			
RN	{1}			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=RF-1;			
RX	MEDLINE-99231861; PubMed-10217509;			
RA	Huang T.C., Jin R.F., Chu M.K., Chen H.M.;			
RT	"Organization and expression of nitrogen-fixation genes in the aerobic			
RT	nitrogen-fixing unicellular cyanobacterium Synechococcus sp. strain			
RF	RF-1.";			
RL	Microbiology 145:743-753(1999).			
DR	EMBL AF003700; AAC5193.1; -.			
FT	NON-TER	1		
SO	SEQUENCE	7 AA;	849 MW;	7412C72AA9D5E030 CFC6;

Query Match	28.1%	Score 9;	DB 2;	Length 7;
Best Local Similarity	66.7%	Pred. No.	5.6e+05;	
Matches	2;	Conservative	0;	Mismatches 1;
				Indels 0;
				Gaps 0
Oy	1 FTL 3			
	1			
	1			
Db	3 FDL 5			
RESULT	8			
P82445				
ID	P82445	PRELIMINARY;	PRT;	7 AA.
AC	P82445;			
DT	01-JUN-2000 (TREMBlrel. 14, Created)			
DT	01-JUN-2000 (TREMBlrel. 14, Last sequence update)			
DT	01-JUN-2000 (TREMBlrel. 14, Last annotation update)			

DE 10 KDA EEL WALL PROTEIN (FRAGMENT).  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota: Viridiplantae: Streptophyta: Tracheophyta  
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:  
OC Asteridae: euasterids I: Solanales: Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP  
RC SEQUENCE.  
RC SPRAIN-CV. PETIT HAVANA.  
RA Blee K A., Bonham V.A., Mitchell G.P., Robertson D., Stabas A.R.,  
RA Moltiszek P. Boljvell G.P.;  
RT "Proteomic study of secondary cell wall proteins from transformed  
RT tobacco cultures".  
RL Plants 0:0-0(2000).  
CC -1- SUBCELLULAR LOCATION: CELL WALL.  
CC -1- TISSUE SPECIFICITY: XYLEM.  
KW Cell wall.  
FT NON\_PTM 7  
SQ SEQUENCE 7 AA; 755 MM; 69D2C1E862D1B2A0 CRC64;

```
OY      3 LTI 5  
        :|:  
DB       1 VIV 3
```

Query Match            28.1%; Score 9; DB 10; Length 7;  
Best Local Similarity 33.3%  
Matches         1; Conservative      2; Mismatches    0; Indels        0; Gaps          0;

RESULT	9	
P82181		
ID	P82181	PRELIMINARY; PRT; 6 AA.
AC	P82181;	
DT	01-JUN-2000 (TREMBLrel. 14, Created)	
DT	01-JUN-2000 (TREMBLrel. 14, Last sequence update)	
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE	CHLOROPLAST 50S RIBOSOMAL PROTEIN L10 BETA (FRAGMENT).	
OS	Splachna olereacea (Splachn).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	
OC	Caryophyllales; Caryophyllales; Chenopodiaceae; Splachna.	
OX	NCBI_Taxid=3562;	
RY	[1]	
RP	SEQUENCE.	
RC	STRAIN-CV. ALMARO; TISSUE=LEAF;	
RX	MEDLINE=20435798; PubMed=10874046;	
RA	Yamaguchi K., Subramanian A.R.;	
RT	"The plastid ribosomal proteins: identification of all the proteins in	
RI	the 50 S subunit of an organelle ribosome (chloroplast).";	
RL	J. Biol. Chem. 275:28466-28482(2000).	
CC	-1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.	
CC	-1- SUBCELLULAR LOCATION: CHLOROPLAST.	
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES.	
CC	-1- MISCELLANEOUS: ON THE 2D-GEL ITS MW IS: 16.5 KDA.	
CC	-1- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.	
DR	Interpro: IPR001790; Ribosomal_L10.	
DR	Interpro: IPR002363; Ribosomal_L10_eub.	
DR	Pfam: PF00466; Ribosomal_L10_PARTIAL.	
DR	PROSITE: PS01109; RIBOSOMAL_L10_PARTIAL.	
RW	Ribosomal protein; Chloroplast; rRNA-binding.	
FT	NON_TER	
SO	SEQUENCE	
	6 AA; 675 MW; 6321B415B05DB000 CRC64;	

Query Match	25.0%;	Score 8;	DB 10;	Length 6;
Best Local	Similarity	100.0%;	pred. No. 5.6e+05;	
Matches	2;	Conservative	0;	Mismatches 0;
				Indels 0;
Gy	5 IS	6		
	/	/		
Db	2 IS	3		

```

RESULT 10
ID P82182; PRELIMINARY; PRT; 6 AA.
AC P82182;
DT 01-JUN-2000 (TREMBLrel. 14, Created)
DT 01-JUN-2000 (TREMBLrel. 14, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CHLOROPLAST 50S RIBOSOMAL PROTEIN L10 GAMMA (FRAGMENT).
OS Spinacia oleracea (Spinach).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicotyledons:
OC Caryophyllales: Caryophyllales: Chenopodiaceae: Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE.
RC STRAIN=CV. ALVARO; TISSUE=LEAF;
RX MEDLINE=20435798; PubMed=10874046;
RA Yamaguchi K., Subramanian A.R.;
RT "The plastid ribosomal proteins. Identification of all the proteins in
RT the 50 S subunit of an organelle ribosome (chloroplast).";
RL J. Biol. Chem. 275:28466-28482(2000)
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.
CC -1- SUPPLEMENTARY LOCATION: CHLOROPLAST.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES.
CC -1- MISCELLANEOUS: ON THE 2D-GEL ITS MW IS: 16.5 KDA.
CC -1- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
DR InterPro: IPR001790; RIBOSOMAL_L10.
DR Pfam: PF00466; RIBOSOMAL_L10; PARTIAL.
DR PROSITE: PS01109; RIBOSOMAL_L10; PARTIAL.
KM Ribosomal protein; Chloroplast; rRNA-binding.
FT NON_TER 1
SQ SEQUENCE 6 AA: 675 MW: 6321B415B05DB000 CRC64;

Query Match 25.0%; Score 8; DB 10; Length 6;
Best Local Similarity 100.0%; Pred. No. 5.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 IS 6
DB 2 IS 3

RESULT 11
ID P72081; PRELIMINARY; PRT; 7 AA.
AC P72081;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 3'-METHYLCEPHEM HYDROXYLASE (FRAGMENT).
OS CEPE.
OC Nocardia lactamdurans.
OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:
OC Actinomycetales: Pseudonocardiaceae: Pseudonocardiaceae: Amycolatopsis.
OX NCBI_TaxID=1913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96009872; PubMed=7557411;
RA Coque J., Perez-Llarena F.J., Enguita F.J., Fuente J.L., Martin J.F.,
RA Liras P.;
RT "Characterization of the cmcH genes of Nocardia lactamdurans and
RT Streptomyces clavuligerus encoding a functional 3'-hydroxymethylcephem
RT O-carbamoyltransferase for cephamycin biosynthesis.";
RL Gene 162:21-27(1995).
DR EMBL: Z21682; CAA79797.1; -.
FT NON_TER 1
SQ SEQUENCE 7 AA: 746 MW: 75A1B2CDD1E681C0 CRC64;

Query Match 25.0%; Score 8; DB 2; Length 7;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Best Local Similarity 33.3%; Pred. No. 5.6e+05;
Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 5 ISS 7
DB 5 VTS 7

RESULT 12
ID 047029; PRELIMINARY; PRT; 7 AA.
AC 047029;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE AAD A1 PROTEIN (FRAGMENT).
GN AAD A1.
OS Enterobacter cloacae.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae;
OC Enterobacter.
OX NCBI_TaxID=550;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94079349; PubMed=8257126;
RA Rather P.N., Mann P.A., Mierzwa R., Hare R.S., Miller G.H., Shaw K.J.;
RT "Analysis of the aac(3)-Via gene encoding a novel 3-N-
RT acetyltransferase.";
RL Antimicrob. Agents Chemother. 37:2074-2079(1993).
DR EMBL: M88012; AAA16193.1; -.
FT NON_TER 1
SQ SEQUENCE 7 AA: 744 MW: 633662DC231A030 CRC64;

Query Match 21.9%; Score 7; DB 2; Length 7;
Best Local Similarity 50.0%; Pred. No. 5.6e+05;
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 LT 4
DB 1 IT 2

RESULT 13
ID 065578; PRELIMINARY; PRT; 7 AA.
AC 065578;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEPHERETICAL 0.8 KDA PROTEIN (FRAGMENT).
OS bovine herpesvirus 1.
OC Viruses; dsDNA viruses, no RNA stage: Herpesviridae;
OC Alphaherpesvirinae: Varicelloviruses.
OX NCBI_TaxID=10320;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=COOPER.
RX MEDLINE=95313343; PubMed=7793062;
RA Vitek C., Benes V., Lu Z., Kutlish G.F., Paces V., Rock D.,
RA Lachnawich G.J., Schwyzer M.;
RT "Nucleotide sequence analysis of a 30-kb region of the bovine
RT herpesvirus 1 genome which exhibits a colinear gene arrangement with
RT the UL21 to UL4 genes of herpes simplex virus.";
RL Virology 210:100-108(1995).
DR EMBL: Z48053; CAA86130.1; -.
KM Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 7 AA: 758 MW: 6DD33455B1F1D1C0 CRC64;

Query Match 21.9%; Score 7; DB 12; Length 7;
Best Local Similarity 50.0%; Pred. No. 5.6e+05;
Matches 2; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 4 TISS 7  
1  
DB 1 TTHS 4

## RESULT 14

ID P82070 PRELIMINARY; PRT: 5 AA.  
AC P82070;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
DE RUBELLIDIN 1.1  
OS Litoria rubella (Desert tree frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litoria.  
OX NCBI\_TaxID=104895;  
RN [1]  
RP SEQUENCE, AND MASS SPECTROMETRY.  
RC TISSUE-SKIN SECRETION;  
RA Steinhorn S.T., Wabnitz P.A., Maugh R.J., Bowie J.H., Gao C.,  
RA Tyler M.J., Wallace J.C.;  
RT \*The structure of new peptides from the Australian red tree frog  
RT 'Litoria rubella', the skin peptide profile as a probe for the study  
RT of evolutionary trends of amphibians.";  
RL Aust. J. Chem. 49:955-963(1996).  
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR  
CC ANTI-BIOTIC ACTIVITY.  
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
CC -1- MASS SPECTROMETRY: MW=598; METHOD=FAb.  
KW Amphibian skin.  
SQ SEQUENCE 5 AA: 598 MW: 60D9C9CAB2A00000 CRC64;

Query Match 18.8%; Score 6; DB 13; Length 5;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 F 1  
1  
DB 3 F 3

## RESULT 15

ID P82071 PRELIMINARY; PRT: 5 AA.  
AC P82071;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
DE RUBELLIDIN 2.1  
OS Litoria rubella (Desert tree frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litoria.  
OX NCBI\_TaxID=104895;  
RN [1]  
RP SEQUENCE, AND MASS SPECTROMETRY.  
RC TISSUE-SKIN SECRETION;  
RA Steinhorn S.T., Wabnitz P.A., Maugh R.J., Bowie J.H., Gao C.,  
RA Tyler M.J., Wallace J.C.;  
RT \*The structure of new peptides from the Australian red tree frog  
RT 'Litoria rubella', the skin peptide profile as a probe for the study  
RT of evolutionary trends of amphibians.";  
RL Aust. J. Chem. 49:955-963(1996).  
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR  
CC ANTI-BIOTIC ACTIVITY.  
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
CC -1- MASS SPECTROMETRY: MW=626; METHOD=FAb.  
KW Amphibian skin.  
SQ SEQUENCE 5 AA: 626 MW: 6DD9C9C9B10300000 CRC64;

Query Match 18.8%; Score 6; DB 13; Length 5;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 F 1  
1  
DB 3 F 3

Search completed: July 15, 2002, 13:37:39  
Job time: 729 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:01:09 ; Search time 107.71 seconds

(without alignments)  
7.219 Million cell updates/sec

Title: US-09-712-819A-1

Perfect score: 32

Sequence: 1 FRLTSS 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 52936

Minimum DB seq length: 0

Maximum DB seq length: 7

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: A.Geneseq.032802.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	87.5	7	20	AAV40738
2	28	87.5	7	21	AAV40738
3	25	78.1	7	20	AAV40737
4	25	78.1	7	21	AAV40737
5	24	75.0	7	20	AAV40736
6	24	75.0	7	21	AAV40736
7	21	65.6	7	20	AAV40735
8	21	65.6	7	21	AAV40735
9	19	59.4	6	22	AAV51422
10	18	56.2	7	19	AAV56711
11	18	56.2	7	21	AAV64225

12	17	53.1	6	18	AAV39453
13	17	53.1	6	19	AAV75358
14	17	53.1	6	19	AAV75290
15	17	53.1	7	11	AAV07656
16	17	53.1	7	20	AAV40723
17	17	53.1	7	21	AAV30062
18	17	53.1	7	22	AAV44200
19	17	53.1	7	22	AAV44205
20	16	50.0	6	10	AAV93345
21	16	50.0	6	17	AAV93089
22	16	50.0	6	19	AAV61443
23	16	50.0	6	20	AAV92487
24	16	50.0	6	21	AAV03571
25	16	50.0	6	21	AAV90491
26	16	50.0	6	22	AAV60836
27	16	50.0	6	22	AAV36961
28	16	50.0	6	22	AAV72775
29	16	50.0	7	20	AAV42013
30	16	50.0	7	20	AAV40625
31	16	50.0	7	20	AAV13747
32	16	50.0	7	21	AAV29964
33	16	50.0	7	21	AAV83858
34	16	50.0	7	22	AAV44336
35	16	50.0	7	22	AAV44453
36	16	50.0	7	22	AAV44458
37	16	50.0	7	22	AAV44632
38	16	50.0	7	22	AAV46017
39	16	50.0	7	22	AAV46022
40	16	50.0	7	22	AAV46032
41	16	50.0	7	22	AAV46037
42	16	50.0	7	22	AAV46042
43	16	50.0	7	22	AAV46044
44	16	50.0	7	22	AAV46044
45	16	50.0	7	22	AAV46044

## ALIGNMENTS

RESULT 1	
AAV40738	AAV40738 standard; peptide; 7 AA.
AAV40738:	
01-DEC-1999 (first entry)	
S4 derivative #12, beta strand of scaffold protein structure.	
Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;	
tumour; chemotherapeutic agent.	
Synthetic.	
EP947582-A1.	
06-OCT-1999.	
31-MAR-1998; 98EP-0870065.	
31-MAR-1998; 98EP-0870065.	
(INNO-) INNOGENETICS NV.	
Desmet J, Hufton S, Hoogenboom H, Sablon E;	
WPI; 1999-542958/46.	
New scaffold protein, useful for stabilizing antigens used as vaccines	
-	
Disclosure; Page 6; 105pp; English.	

CC Sequences AAY40727-Y40748 are functionally equivalent derivatives of the  
CC S4 peptide (AAY40607) which forms part of a scaffold protein. S4 is a  
CC beta strand peptide which forms part of a beta sheet. Peptides  
CC (AAY40601-Y40609) together form a single-chain scaffold protein which  
CC contains at least 1 disulfide bond, contains less than 10% alpha helix  
CC and contains at least 6 beta-strands. The scaffold protein is constructed  
CC of beta strands S1-S6, and may also include beta strands A1-A3, or any  
CC functionally equivalent derivative of these sequences. The beta strands  
CC form two beta sheets S1/S4/S3 and S6/S5/S2 with each strand connected to  
CC the next by hydrogen bonds, which generate a beta sandwich architecture.  
CC If the additional beta strands A1-A3 are included in the structure the  
CC scaffold is constructed of two beta sheets, with the structures  
CC A1/S1/S4/S3 and S6/S5/S2/A2/A3. The beta strands are connected to each  
CC other via amino acid loops, where at least one of the loops binds to a  
CC receptor or antigen. The scaffold protein is used to stabilize antigens  
CC or whole proteins such as receptors, or their fragments. It may be used  
CC to bind two separate molecules. For example, one surface of the scaffold  
CC may be bound to a protein which binds to a tumour antigen. This will  
CC target the complex to tumour cells. Another surface may be bound to a  
CC cytotoxic molecule or an autoimmune antibody which may then kill the  
CC tumour cells. Therefore the scaffold protein may be used to target  
CC chemotherapeutic agents to specific cells. It may also be used to  
CC stabilize individual peptides in a peptide library and may be used in  
CC diagnostic techniques, and to stabilize antigens used as vaccines.

SO Sequence 7 AA;

Query Match 87.5%; Score 28; DB 20; Length 7;  
Best Local Similarity 100.0%; Pred. No. 6,4e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTIS 6  
| | | | |  
Db 2 flltis 7

RESULT 2  
AAB30076  
ID AAB30076 standard; Peptide: 7 AA.  
XX  
AC AAB30076;  
XX  
DT 09-FEB-2001 (first entry)  
XX  
DE Scaffold protein SCA S4 peptide SEQ ID NO: 137.  
XX  
KW Human; CTLA-4; scaffold protein; antigen-binding; receptor-binding;  
KW SCA domain; cancer; thrombosis; osteoporosis; rheumatoid arthritis;  
KW diabetic retinopathy; atherosclerosis.  
XX  
OS Synthetic.  
XX  
PM WO200060070-A1.  
XX  
PD 12-OCT-2000.  
XX  
PF 01-APR-1999; 99WO-EP02283.  
XX  
PR 01-APR-1999; 99WO-EP02283.  
XX  
PA (INNO-) INNOGENETICS NV.  
XX  
PI Desmet J, Hufton S, Hoogenboom H, Sablon E;  
XX  
DR WPI: 2000-665002/64.  
XX  
PT Scaffold composed of single-chain polypeptide having beta sandwich  
PT architecture carrying new and randomized peptide sequences useful as  
PT supporting framework and carrying antigen- or receptor binding  
PT fragments  
XX  
PS Disclosure: Page 15; 68pp; English.

XX  
CC The present invention is concerned with producing scaffold proteins  
CC based upon the human CTLA-4 SCA domain. These scaffold proteins can be  
CC used as a scaffold to bind antigen- or receptor-binding fragments. These  
CC can be used in the treatment of diseases such as cancer,  
CC atherosclerosis, thrombosis, osteoporosis, rheumatoid arthritis and  
CC diabetic retinopathy. Sequences AAB29930-A29939 were used in the  
CC production of the proteins of the invention.

SO Sequence 7 AA;

Query Match 87.5%; Score 28; DB 21; Length 7;  
Best Local Similarity 100.0%; Pred. No. 6,4e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTIS 6  
| | | | |  
Db 2 flltis 7

RESULT 3  
AAY40737  
ID AAY40737 standard; peptide: 7 AA.  
XX  
AC AAY40737;  
XX  
DT 01-DEC-1999 (first entry)  
XX  
DE S4 derivative #11, beta strand of scaffold protein structure.  
XX  
KW Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;  
KW tumour; chemotherapeutic agent.  
XX  
OS Synthetic.  
XX  
PN EP947582-A1.  
XX  
PD 06-OCT-1999.  
XX  
PF 31-MAR-1998; 98EP-0870065.  
XX  
PR 31-MAR-1998; 98EP-0870065.  
XX  
PA (INNO-) INNOGENETICS NV.  
XX  
PI Desmet J, Hufton S, Hoogenboom H, Sablon E;  
XX  
DR WPI: 1999-542958/46.  
XX  
PT New scaffold protein, useful for stabilizing antigens used as vaccines  
PT  
PS Disclosure: Page 6; 105pp; English.  
XX  
XX  
CC Sequences AAY40727-Y40748 are functionally equivalent derivatives of the  
CC S4 peptide (AAY40607) which forms part of a scaffold protein. S4 is a  
CC beta strand peptide which forms part of a beta sheet. Peptides  
CC (AAY40601-Y40609) together form a single-chain scaffold protein which  
CC contains at least 1 disulfide bond, contains less than 10% alpha helix  
CC and contains at least 6 beta-strands. The scaffold protein is constructed  
CC of beta strands S1-S6, and may also include beta strands A1-A3, or any  
CC functionally equivalent derivative of these sequences. The beta strands  
CC form two beta sheets S1/S4/S3 and S6/S5/S2 with each strand connected to  
CC the next by hydrogen bonds, which generate a beta sandwich architecture.  
CC If the additional beta strands A1-A3 are included in the structure the  
CC scaffold is constructed of two beta sheets, with the structures  
CC A1/S1/S4/S3 and S6/S5/S2/A2/A3. The beta strands are connected to each  
CC other via amino acid loops, where at least one of the loops binds to a  
CC receptor or antigen. The scaffold protein is used to stabilize antigens  
CC or whole proteins such as receptors, or their fragments. It may be used  
CC to bind two separate molecules. For example, one surface of the scaffold  
CC may be bound to a protein which binds to a tumour antigen. This will



CC target the complex to tumour cells. Another surface may be bound to a  
 CC cytotoxic molecule or an autoimmune antibody which may then kill the  
 CC tumour cells. Therefore the scaffold protein may be used to target  
 CC chemotherapeutic agents to specific cells. It may also be used to  
 CC stabilize individual peptides in a peptide library and may be used in  
 CC diagnostic techniques, and to stabilize antigens used as vaccines.  
 XX

SO Sequence 7 AA:

Query Match 78.18; Score 25; DB 20; Length 7;  
 Best Local Similarity 83.3%; Pred. No. 6.4e+05;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTIS 6  
 :|||||  
 Db 2 ytlfts 7

RESULT 4

AAB30075 ID AAB30075 standard; Peptide; 7 AA.

AC AAB30075;

DT 09-FEB-2001 (first entry)

DE Scaffold protein SCA S4 peptide SEQ ID NO: 136.

KW Human; CTLA-4; scaffold protein; antigen-binding; receptor-binding;  
 KW SCA domain; cancer; thrombosis; osteoporosis, rheumatoid arthritis;  
 KW diabetic retinopathy; atherosclerosis.

OS Synthetic.

PN MO200060070-A1.

PD 12-OCT-2000.

PF 01-APR-1999; 99MO-EP02283.

PR 01-APR-1999; 99MO-EP02283.

RA (INNO-) INNOGENETICS NV.

PJ Desmet J, Hufton S, Hoogenboom H, Sablon E;

DR WPI; 2000-663002/64.

XX Scaffold composed of single-chain polypeptide having beta sandwich  
 PT architecture carrying new and randomized peptide sequences useful as  
 PT supporting framework and carrying antigen- or receptor binding  
 PT fragments -

PS Disclosure; Page 15; 68pp; English.

XX The present invention is concerned with producing scaffold proteins  
 CC based upon the human CTLA-4 SCA domain. These scaffold proteins can be  
 CC used as a scaffold to bind antigen- or receptor-binding fragments. These  
 CC can be used in the treatment of diseases such as cancer,  
 CC atherosclerosis, thrombosis, osteoporosis, rheumatoid arthritis and  
 CC diabetic retinopathy. Sequences AAB2930-B2933 were used in the  
 CC production of the proteins of the invention.  
 CC

SO Sequence 7 AA:

Query Match 78.18; Score 25; DB 21; Length 7;  
 Best Local Similarity 83.3%; Pred. No. 6.4e+05;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTIS 6  
 :|||||

Db 2 ytlfts 7

RESULT 5

AAV40736 ID AAV40736 standard; peptide; 7 AA.

AC AAV40736;

DT 01-DEC-1999 (first entry)

DE S4 derivative #10, beta strand of scaffold protein structure.

KW Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;  
 KW tumour; chemotherapeutic agent.

OS Synthetic.

PN EP947582-A1.

PD 06-OCT-1999.

PF 31-MAR-1998; 98EP-0870065.

PR 31-MAR-1998; 98EP-0870065.

RA (INNO-) INNOGENETICS NV.

PJ Desmet J, Hufton S, Hoogenboom H, Sablon E;

DR WPI; 1999-542958/46.

DE New scaffold protein, useful for stabilizing antigens used as vaccines

PS Disclosure; Page 6; 105pp; English.

XX Sequences AAV40727-Y40748 are functionally equivalent derivatives of the  
 CC S4 peptide (AAV40607) which forms part of a scaffold protein. S4 is a  
 CC beta strand peptide which forms part of a beta sheet. Peptides  
 CC (AAV40601-Y40609) together form a single-chain scaffold protein which  
 CC contains at least 1 disulfide bond, contains less than 10% alpha helix  
 CC and contains at least 6 beta-strands. The scaffold protein is constructed  
 CC of beta strands S1-S6, and may also include beta strands A1-A3, or any  
 CC functionally equivalent derivative of these sequences. The beta strands  
 CC form two beta sheets S1/S4/S3 and S6/S5/S2 with each strand connected to  
 CC the next by hydrogen bonds, which generate a beta sandwich architecture.  
 CC If the additional beta strands A1-A3 are included in the structure the  
 CC scaffold is constructed of two beta sheets, with the structures  
 CC A1/S1/S4/S3 and S6/S5/S2/A2/A3. The beta strands are connected to each  
 CC other via amino acid loops, where at least one of the loops binds to a  
 CC receptor or antigen. The scaffold protein is used to stabilize antigens  
 CC or whole proteins such as receptors, or their fragments. It may be used  
 CC to bind two separate molecules. For example, one surface of the scaffold  
 CC may be bound to a protein which binds to a tumour antigen. This will  
 CC target the complex to tumour cells. Another surface may be bound to a  
 CC cytotoxic molecule or an autoimmune antibody which may then kill the  
 CC tumour cells. Therefore the scaffold protein may be used to target  
 CC chemotherapeutic agents to specific cells. It may also be used to  
 CC stabilize individual peptides in a peptide library and may be used in  
 CC diagnostic techniques, and to stabilize antigens used as vaccines.  
 CC

SO Sequence 7 AA:

Query Match 75.08; Score 24; DB 20; Length 7;  
 Best Local Similarity 83.3%; Pred. No. 6.4e+05;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTIS 6  
 :|||||  
 Db 2 ftlfts 7

RESULT 6  
AAB30074  
ID AAB30074 standard; Peptide: 7 AA.  
XX  
AC AAB30074;  
XX  
XX  
DT 09-FEB-2001 (first entry)  
XX  
DE Scaffold protein SCA S4 peptide SEQ ID NO: 135.  
XX  
XX Human; CTA-4; scaffold protein; antigen-binding; receptor-binding;  
KW SCA domain; cancer; thrombosis; osteoporosis; rheumatoid arthritis;  
KM diabetic retinopathy; atherosclerosis.  
XX  
OS Synthetic.  
XX  
PN WO200060070-A1.  
XX  
PD 12-OCT-2000.  
XX  
PF 01-APR-1999; 99NO-EP02283.  
XX  
PR 01-APR-1999; 99NO-EP02283.  
XX  
PA (INNO-) INNOGENETICS NV.  
XX  
PI Desmet J, Hufton S, Hoogenboom H, Sablon E;  
XX WPI: 2000-665002/64.  
DR  
XX  
PT Scaffold composed of single-chain polypeptide having beta sandwich  
XX architecture carrying new and randomized peptide sequences useful as  
XX supporting framework and carrying antigen- or receptor binding  
XX fragments -  
XX  
PS Disclosure: Page 15; 68pp; English.  
XX  
CC The present invention is concerned with producing scaffold proteins  
CC based upon the human CTA-4 SCA domain. These scaffold proteins can be  
CC used as a scaffold to bind antigen- or receptor-binding fragments. These  
CC can be used in the treatment of diseases such as cancer,  
CC atherosclerosis, thrombosis, osteoporosis, rheumatoid arthritis and  
CC diabetic retinopathy. Sequences AAB2930-B2939 were used in the  
CC production of the proteins of the invention.  
XX  
SQ Sequence 7 AA:  
  
Query Match 75.0%; Score 24; DB 21; Length 7;  
Best Local Similarity 83.3%; Pred. No. 6.4e+05;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 FRTGIS 6  
DB 2 YELTIS 7  
  
RESULT 7  
AAI40735  
ID AAI40735 standard; peptide: 7 AA.  
XX  
AC AAI40735;  
XX  
XX  
DT 01-DEC-1999 (first entry)  
XX  
DE S4 derivative #9, beta strand of scaffold protein structure.  
XX  
KW Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;  
KM tumour; chemotherapeutic agent.  
XX  
OS Synthetic.  
XX  
PN WO200060070-A1.

PN EP947582-A1.  
XX  
XX 06-OCT-1999.  
PD  
XX  
PE 31-MAR-1998; 98EP-0870065.  
XX  
PR 31-MAR-1998; 98EP-0870065.  
XX  
PA (INNO-) INNOGENETICS NV.  
XX  
PI Desmet J, Hufton S, Hoogenboom H, Sablon E;  
XX WPI: 1999-542958/46.  
XX  
PT New scaffold protein, useful for stabilizing antigens used as vaccines  
XX  
XX  
PS Disclosure: Page 6; 105pp; English.  
XX  
XX  
CC Sequences AAI40727-Y40748 are functionally equivalent derivatives of the  
CC S4 peptide (AAI40607) which forms part of a scaffold protein. S4 is a  
CC beta strand peptide which forms part of a beta sheet. Peptide  
CC (AAI40601-Y40605) together form a single-chain scaffold protein which  
CC contains at least 1 disulfide bond, contains less than 10% alpha helix  
CC and contains at least 6 beta-strands. The scaffold protein is constructed  
CC of beta strands S1-S6, and may also include beta strands A1-A3, or any  
CC functionally equivalent derivative of these sequences. The beta strands  
CC form two beta sheets S1/S4/S5 and S6/S5/S2 with each strand connected to  
CC If the additional beta strands A1-A3 are included in the structure the  
CC scaffold is constructed of two beta sheets, with the structures  
CC A1/S1/S4/S3 and S6/S5/S2/A2/A3. The beta strands are connected to each  
CC other via amino acid loops, where at least one of the loops binds to a  
CC receptor or antigen. The scaffold protein is used to stabilize antigens  
CC or whole proteins such as receptors, or their fragments. It may be used  
CC to bind two separate molecules. For example, one surface of the scaffold  
CC may be bound to a protein which binds to a tumour antigen. This will  
CC target the complex to tumour cells. Another surface may be bound to a  
CC cytotoxic molecule or an autoimmune antibody which may then kill the  
CC tumour cells. Therefore the scaffold protein may be used to target  
CC chemotherapeutic agents to specific cells. It may also be used to  
CC stabilize individual peptides in a peptide library and may be used in  
CC diagnostic techniques, and to stabilize antigens used as vaccines.  
XX  
SQ Sequence 7 AA:  
  
Query Match 65.6%; Score 21; DB 20; Length 7;  
Best Local Similarity 66.7%; Pred. No. 6.4e+05;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 FRTGIS 6  
DB 2 YELTIS 7  
  
RESULT 8  
AAB30073  
ID AAB30073 standard; Peptide: 7 AA.  
XX  
AC AAB30073;  
XX  
XX  
DT 09-FEB-2001 (first entry)  
XX  
DE Scaffold protein SCA S4 peptide SEQ ID NO: 134.  
XX  
XX Human; CTA-4; scaffold protein; antigen-binding; receptor-binding;  
KW SCA domain; cancer; thrombosis; osteoporosis; rheumatoid arthritis;  
KM diabetic retinopathy; atherosclerosis.  
XX  
OS Synthetic.  
XX  
PN WO200060070-A1.

XX 12-OCT-2000.  
PD  
XX 01-APR-1999; 99WO-EP02283.  
XX  
XX 01-APR-1999; 99WO-EP02283.  
XX  
XX (INNO-) INNOGENETICS NV.  
XX  
XX Desmet J, Hufton S, Hoogenboom H, Sablon E;  
XX WPI: 2000-665002/64.  
XX  
XX Scaffold composed of single-chain polypeptide having beta sandwich  
XX architecture carrying new and randomized peptide sequences useful as  
XX supporting framework and carrying antigen- or receptor binding  
XX fragments -  
XX  
XX Disclosure: Page 15; 68pp; English.  
XX  
XX The present invention is concerned with producing scaffold proteins  
XX based upon the human CTLA-4 SCA domain. These scaffold proteins can be  
XX used as a scaffold to bind antigen- or receptor-binding fragments. These  
XX can be used in the treatment of diseases such as cancer,  
XX atherosclerosis, thrombosis, osteoporosis, rheumatoid arthritis and  
XX diabetic retinopathy. Sequences AAB29930-B29939 were used in the  
XX production of the proteins of the invention.  
XX  
XX Sequence 7 AA:  
SQ  
Query Match 65.6%; Score 21; DB 21; Length 7;  
Best Local Similarity 66.7%; Pred. No. 6.4e+05;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FTLTIS 6  
: : : : :  
DB 2 ystlts 7  
RESULT 9  
AAM51422  
ID AAM51422 standard; peptide; 6 AA.  
XX  
XX AAM51422:  
AC 08-JAN-2002 (first entry)  
XX  
XX Integrin activating peptide SEQ ID NO 1.  
DE  
XX Integrin stimulant; vulnerary; injury healing;  
KW postsurgical tissue recovery.  
XX  
XX Unidentified.  
OS  
XX JP2001213898-A.  
PN  
XX  
XX 07-AUG-2001.  
PD  
XX 31-JAN-2000; 2000JP-0022469.  
XX  
XX 31-JAN-2000; 2000JP-0022469.  
PR  
XX (HISM ) HISMITSU PHARM CO LTD.  
XX  
XX WPI: 2001-629610/73.  
DR  
XX An injury healing and postsurgical tissue recovering integrin  
XX activating peptide -  
XX  
XX Claim 1; Page 3; 11pp; Japanese.  
PS  
XX The invention relates to novel peptides with vulnerary activity, useful

CC for injury healing and postsurgical tissue recovery by acting as an  
CC integrin stimulant.  
XX  
XX Sequence 6 AA:  
SQ  
Query Match 59.4%; Score 19; DB 22; Length 6;  
Best Local Similarity 60.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FTLTIS 5  
: : : : :  
DB 1 ystlts 5  
RESULT 10  
AAM58711  
ID AAM58711 standard; peptide; 7 AA.  
XX  
XX AAM58711:  
AC 17-SEP-1998 (first entry)  
XX  
XX Tryptic 40 kD subunit CLMF peptide off PVDF #4.  
DE  
XX Cytotoxic lymphocyte maturation factor; CLMF; cytokine; synergize;  
XX Interleukin 2; IL-2; human; lymphoblastoid cell; detection; purifying;  
XX Proliferation; cytotoxic T cell; transplantation; antibody.  
XX  
XX Homo sapiens.  
OS  
XX US5780597-A.  
PN  
XX 14-JUL-1998.  
PD  
XX 02-JUN-1995; 95US-0460061.  
PE  
XX 24-MAR-1992; 92US-0857023.  
PR 22-DEC-1989; 89US-0455708.  
PR 09-MAY-1990; 90US-0520935.  
PR 27-AUG-1990; 90US-0572284.  
PR 02-MAR-1994; 94US-0205011.  
PR 02-JUN-1995; 95US-0460061.  
XX  
XX (HOFF ) HOFFMANN LA ROCHE INC.  
XX  
XX Chizzonite RA, Gately MK, Gubler UA, Hultes JD;  
XX Pan YE, Podlaski FJ, Stern AS;  
XX WPI: 1998-413150/35.  
DR  
XX  
XX New antibodies to cytotoxic lymphocyte maturation factor - useful  
XX for detecting, purifying, and/or blocking proliferation and  
XX activation of cytotoxic T cells, such as in transplantation(s)  
XX  
XX Example 3; Column 23; 71pp; English.  
XX  
XX An isolated antibody has been developed which binds specifically to  
XX cytotoxic lymphocyte maturation factor (CLMF) and neutralises CLMF  
XX induced proliferation by more than 50% in a CLMF dependent T cell  
XX growth assay and/or inhibits binding of at least 60% of the factor  
XX to phytohemagglutinin (PHA) activated peripheral blood lymphocyte  
XX (PBL) blasts as determined in a CLMF receptor binding assay. CLMF is  
XX a heterodimeric protein having a molecular weight band of 75 kD,  
XX determined by sodium dodecyl sulphate-polyacrylamide gel electrophoresis  
XX (SDS-PAGE) under non-reducing and/or reducing conditions, providing a  
XX first subunit having a molecular weight band of 40 kD and a second  
XX subunit having a molecular weight of 35 kD. The present sequence  
XX represents a tryptic 40 kD CLMF peptide off polyvinylidene difluoride  
XX (PVDF), from an example of the present invention. The antibody can be  
XX used for the purification and/or detection of CLMF. It is also used in  
XX therapeutic treatments which require selective blocking of proliferation  
XX and activation of cytotoxic T cells (CTLs) such as in transplantation.

```

XX Sequence 7 AA:
SQ
Query Match 56.2%; Score 18; DB 19; Length 7;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 TLT1 5
1111
DB 1 tlt1 4

RESULT 11
AAV64225
ID AAV64225 standard; Peptide; 7 AA.
XX
AC AAV64225;
XX
DT 02-MAR-2000 (first entry)
XX
DE Cadherin-related neuronal receptor CAR sequence SEQ ID NO:637.
XX
KM Modulation; nonclassical cadherin mediated cell adhesion; CAR;
KM Inhibition; cadherin extracellular domain; cell adhesion recognition;
KM OB-cadherin; cadherin-5; cadherin-6; cadherin-7; cadherin-8;
KM cadherin-12; cadherin-14; cadherin-15; T-cadherin; PB-cadherin;
KM cadherin related neuronal receptor; LI-cadherin; protocadherin;
KM desmoglein; desmocollin; calcium binding; cancer; tumour; obesity;
KM rheumatoid arthritis; multiple sclerosis; diabetes; metastasis;
KM neurological disease.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN MO9957149-A2.
XX
PD 11-NOV-1999.
XX
PE 05-MAY-1999; 99MO-CA00363.
XX
PR 05-MAY-1998; 98US-0073040.
PR 06-NOV-1998; 98US-0187859.
PR 20-JAN-1999; 99US-0234395.
PR 08-MAR-1999; 99US-0264516.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuk OM, Gour BJ, Byers S;
XX
DR WPI: 2000-038791/03.
XX
PT New cadherin modulating agents, used for modulating nonclassical
PT cadherin-mediated functions for treating e.g. cancers, obesity,
PT rheumatoid arthritis, multiple sclerosis, diabetes or a neurological
PT disease -
XX
PS Claim 99; Page 216; 252pp; English.
XX
XX The present invention describes cadherin modulating agents (MA)
CC comprising peptides which comprise a nonclassical cadherin cell adhesion
CC recognition (CAR) sequence. The MA's can be used for modulating
CC nonclassical cadherin-mediated functions. They can be used for e.g.
CC inhibiting adhesion of nonclassical-cadherin expressing cells in a
CC mammal, enhancing delivery of a drug through the skin of a mammal,
CC enhancing delivery of a drug to a tumour in a mammal, treating cancer in
CC a mammal, inhibiting metastasis of a cancer in a mammal, inhibiting
CC angiogenesis in a mammal, inducing apoptosis in a nonclassical cadherin-
CC expressing cell, preventing or treating obesity in a mammal, stimulating
CC blood vessel regression in a mammal, enhancing drug delivery to the
CC central nervous system, treating a demyelinating neurological disease,
CC increasing vasopermeability in a mammal, enhancing adhesion of
CC nonclassical cadherin-expressing cells, inhibiting synaptic stability in

```

```

CC a mammal, or preventing pregnancy in a mammal. They can also be used for
CC e.g. enhancing or directing neurite outgrowth, facilitating wound
CC healing or reducing scar tissue, or enhancing adhesion of foreign tissue
CC in a mammal. They can also be used for treating e.g. psoriasis,
CC arthritis, age-related macular degeneration, multiple sclerosis and
CC diabetes. The products can also be used for detection and diagnosis and
CC in biosensors. AAV60592 to AAV64572 represent specifically claimed
CC peptides, and AAV64573 to AAV64643 and AA233183 to AA233186 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 7 AA:
XX
Query Match 56.2%; Score 18; DB 21; Length 7;
Best Local Similarity 57.1%; Pred. No. 6.4e+05;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 FTLTSS 7
1111
DB 1 ftlts 7

RESULT 12
AAW39453
ID AAW39453 standard; peptide; 6 AA.
XX
AC AAW39453;
XX
DT 11-JUN-1998 (first entry)
XX
DE Human T cell epitope proteolytic cleavage site 4.
XX
KM T cell epitope; immune response; human leukocyte antigen; HLA class I;
KM vaccine; immunogenic; major histocompatibility complex; MHC; B cell;
KM disease; anti-tumour; anti-viral.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN MO9741440-A1.
XX
PD 06-NOV-1997.
XX
PE 28-APR-1997; 97WO-NL00229.
XX
PR 23-DEC-1996; 96EP-0203670.
PR 26-APR-1996; 96EP-0201145.
XX
PA (UYLF-) RIJKSUNIV LEIDEN.
XX
PI (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
DR Kast WM, Melief CJM, Offringa R, Toes REM, Van Der Burg SH;
XX
DR WPI: 1997-549891/50.
XX
PT Method of selecting T cell peptide epitope(s) - by measuring the
PT stability of HLA class I-peptide complexes on intact B cells
XX
PS Example 5; Page 42; 109pp; English.
XX
XX Peptides AAW39430-W39734 are used in a novel method for the selection of
CC immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC method involves the identification of peptide sequences capable of
CC binding to an HLA (human leukocyte antigen) class I molecule and
CC measuring the binding of this epitope peptide to the HLA class I
CC peptide. The stability of binding of the peptide and MHC (major
CC histocompatibility complex) class I molecule is measured on intact human
CC B cells carrying the MHC molecule at their cell surfaces. The method can
CC be used to select peptide epitopes for generating vaccines against a
CC disease associated with the polypeptide, e.g. cancers or AIDS. The
CC peptide epitopes are especially T-cell peptide epitopes with strong
CC anti-tumour and anti-viral immune responses. Peptides AAW39450-W39503 are
CC proteolytic cleavage sites capable of linking together T cell epitopes in

```

```

CC vaccines.
XX
SQ Sequence 6 AA:

Query Match      53.1%; Score 17; DB 18; Length 6;
Best Local Similarity 66.7%; Pred. No. 6.4e+05;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PMLTIS 6
   | | |
Db 1 fclt1s 6

RESULT 13
AAW75358
ID AAW75358 standard; peptide; 6 AA.
AC
XX AAW75358;
XX
DF 02-FEB-1999 (first entry)
DE Hexapeptide #13 binds anti-Ad5 fibre head MAb 7A2.7.
XX
XX Cellular receptor; virus; immobilisation; monoclonal antibody; fibre;
KW viral surface protein; hexapeptide expression library; adenovirus;
KM major histocompatibility complex; MHC; fibronectin; gene therapy;
KW genetic disease; acquired immune deficiency syndrome; AIDS; cancer.
XX
OS Synthetic.
XX
PN FR2758821-A1.
PD 31-JUL-1998.
XX
PF 30-JAN-1997; 97FR-0001005.
XX
PR 30-JAN-1997; 97FR-0001005.
PR 09-SEP-1997; 97FR-0011166.
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Boulanger P, Hong SS, Karayan L;
DR WPI: 1998-416458/36.
XX
PT Use of peptide(s) for facilitating or modulating attachment of
PT adenovirus to cells - useful for, e.g. treating or preventing
PT infection and improving uptake of gene therapy vectors
XX
PS Example 2; Fig 1; 48pp; French.
XX
XX The invention relates to methods for selecting and identifying a
XX cellular receptor for a virus, by immobilising, on a support, a
XX monoclonal antibody targeted to a viral surface protein that determines
XX attachment of the virus to the receptor. The immobilised antibody is
XX incubated with a hexapeptide expression library and peptides bound to the
XX immobilised antibody are eluted by competitive binding with recombinant
XX fragments of the viral surface protein. In a reverse method, the viral
XX surface protein is immobilised and incubated with the peptide library. In
XX this case, the bound peptides are eluted by competitive binding with the
XX monoclonal antibody. The hexapeptides AAW75346-W75360 represent
XX monoclonal antibody (MAb) 7A2.7 and eluted using an adenovirus type 5
XX fibre head protein (AAW75335).
XX The methods are used to identify peptides from MHC Class I and/or
XX type III fibronectin proteins that allow or facilitate attachment by
XX adenovirus (Ad) to host cells and/or entry into the cells, and to
XX identify ligands that modulate Ad infection mediated by these peptides,
XX e.g. to treat or prevent Ad infections or to facilitate infection by Ad
XX gene therapy vectors used to treat genetic diseases, acquired immune
XX deficiency syndrome or cancer.

```

```

SQ Sequence 6 AA:

Query Match      53.1%; Score 17; DB 19; Length 6;
Best Local Similarity 75.0%; Pred. No. 6.4e+05;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMLT 4
   | | |
Db 3 fclt 6

RESULT 14
AAW75290
ID AAW75290 standard; peptide; 6 AA.
AC
XX AAW75290;
XX
DF 02-FEB-1999 (first entry)
DE Hexapeptide #13 binds anti-Ad5 fibre head MAb 7A2.7.
XX
XX Cellular receptor; virus; immobilisation; monoclonal antibody; fibre;
KW viral surface protein; hexapeptide expression library; adenovirus;
KM major histocompatibility complex; MHC; fibronectin; gene therapy;
KW genetic disease; acquired immune deficiency syndrome; AIDS; cancer.
XX
OS Synthetic.
XX
PN FR2758885-A1.
PD 31-JUL-1998.
XX
PF 30-JAN-1997; 97FR-0001005.
XX
PR 30-JAN-1997; 97FR-0001005.
PR 10-JUL-1997; 97FR-0008796.
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Boulanger P, Hong SS, Karayan L;
DR WPI: 1998-416493/36.
XX
PT Selection and identification of cellular receptors for viruses -
PT used to control attachment and entry of adenovirus into cells, e.g.
PT for treating infection or in gene therapy
XX
PS Example 2; Fig 1; 43pp; French.
XX
XX The invention relates to methods for selecting and identifying a
XX cellular receptor for a virus, by immobilising, on a support, a
XX monoclonal antibody targeted to a viral surface protein that determines
XX attachment of the virus to the receptor. The immobilised antibody is
XX incubated with a hexapeptide expression library and peptides bound to the
XX immobilised antibody are eluted by competitive binding with recombinant
XX fragments of the viral surface protein. In a reverse method, the viral
XX surface protein is immobilised and incubated with the peptide library. In
XX this case, the bound peptides are eluted by competitive binding with the
XX monoclonal antibody. The hexapeptides AAW75278-W75292 represent
XX monoclonal antibody (MAb) 7A2.7 and eluted using an adenovirus type 5
XX fibre head protein (AAW75267).
XX The methods are used to identify peptides from MHC Class I and/or
XX type III fibronectin proteins that allow or facilitate attachment by
XX adenovirus (Ad) to host cells and/or entry into the cells, and to
XX identify ligands that modulate Ad infection mediated by these peptides,
XX e.g. to treat or prevent Ad infections or to facilitate infection by Ad
XX gene therapy vectors used to treat genetic diseases, acquired immune
XX deficiency syndrome or cancer.
XX
SQ Sequence 6 AA:

```

Query Match 53.1%; Score 17; DB 19; Length 6;  
 Best Local Similarity 75.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLT 4  
 :|||  
 Db 3 ylit 6

## RESULT 15

AA07656  
 ID AAR07656 standard; protein: 7 AA.

XX  
 AC AAR07656;

XX  
 DT 04-JAN-1991 (first entry)

XX  
 DE Ribonuclease reductase inhibitor #9.

XX  
 KM mammalian ribonuclease reductase inhibitors; tumour treatment;  
 KM psoriasis treatment;

XX  
 OS Synthetic.

XX  
 FH Key Location/Qualifiers  
 FT Modified-site 1:1  
 FT /label=N-acetyl-L-phenylalanine

XX  
 PN EP83190-A.

XX  
 PD 22-AUG-1990.

XX  
 PF 08-FEB-1990; 90EP-0102453.

XX  
 PR 17-FEB-1989; 89CA-2591372.

XX  
 PA (BIOM-) BIO-MEGA INC.

XX  
 PI Guindon Y, Lavellee P, Rahit S, Cosentino GP;

XX  
 DR WPI; 1990-253483/34.

XX  
 PT Oligopeptide derivs. Inhibitors of mammalian ribonucleotide  
 PT reductase - for control of neoplasms, tumours, etc..

XX  
 PS Claim 10; Page 15; 16pp; English.

XX  
 CC The peptide is

XX  
 CC synthesised by standard solid-phase methods. It is a selective  
 CC inhibitor of mammalian ribonucleotide reductase and can be used to

XX  
 CC control abnormal cell proliferation. The inhibitor has relatively low  
 CC toxicity and can be applied topically or systemically.

XX  
 CC See also AAR06487-R06488, AAR07650-R07651 and AAR07657.

XX  
 SQ Sequence 7 AA;

Query Match 53.1%; Score 17; DB 11; Length 7;  
 Best Local Similarity 60.0%; Pred. No. 6.4e+05;

Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTLT 5  
 :|||  
 Db 1 fltdv 5

Search completed: July 15, 2002, 13:25:24  
 Job time: 1455 sec

GenCore version 4.5  
Copyright (C) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:18:00 ; Search time 40.89 Seconds

(without alignments)  
4.181 Million cell updates/sec

Title: US-09-712-819A-1

Perfect score: 32

Sequence: 1 FTITISS 7

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 39160

Minimum DB seq length: 0

Maximum DB seq length: 7

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCOMUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	56.2	7	4	US-09-187-859-637
2	17	53.1	5	3	US-08-757-177-16
3	17	53.1	7	2	US-08-739-401A-6
4	16	50.0	4	4	US-08-860-904-9
5	16	50.0	6	1	US-08-252-9950-7
6	16	50.0	6	2	US-08-478-386A-46
7	16	50.0	6	2	US-08-292-597-46
8	16	50.0	6	2	US-08-388-653-46
9	16	50.0	6	2	US-08-473-985-46
10	16	50.0	6	2	US-08-834-108-7
11	16	50.0	6	2	US-08-483-898-46
12	16	50.0	6	3	US-09-087-716-46
13	16	50.0	6	3	US-09-157-753-46
14	16	50.0	6	3	US-09-157-230-46
15	16	50.0	6	3	US-09-087-811-46
16	16	50.0	6	3	US-09-156-855-46
17	16	50.0	6	4	US-09-158-010-46
18	16	50.0	6	4	US-09-087-647-46
19	16	50.0	6	4	US-09-302-629-46
20	16	50.0	7	1	US-08-136-743B-55
21	16	50.0	7	2	US-08-177-109A-7
22	16	50.0	7	2	US-08-687-706-7
23	16	50.0	7	3	US-09-040-216-28
24	15	46.9	5	1	US-08-136-743B-63
25	15	46.9	5	2	US-08-667-001-22
26	15	46.9	5	3	US-09-040-216-55
27	15	46.9	5	4	US-08-591-632-17

28	15	46.9	5	4	US-08-591-632-23	Sequence 23, Appl
29	15	46.9	5	4	US-08-591-632-26	Sequence 26, Appl
30	15	46.9	6	1	US-08-136-743B-62	Sequence 62, Appl
31	15	46.9	6	1	US-08-297-731-4	Sequence 4, Appl
32	15	46.9	6	1	US-08-290-448A-41	Sequence 41, Appl
33	15	46.9	6	1	US-08-290-448A-41	Sequence 41, Appl
34	15	46.9	6	1	US-08-175-069A-41	Sequence 41, Appl
35	15	46.9	6	2	US-08-622-770A-8	Sequence 8, Appl
36	15	46.9	6	3	US-09-040-216-54	Sequence 54, Appl
37	15	46.9	6	3	US-08-297-395-3	Sequence 3, Appl
38	15	46.9	6	3	US-09-146-675-6	Sequence 6, Appl
39	15	46.9	6	4	US-09-623-618B-6	Sequence 6, Appl
40	15	46.9	6	4	US-08-461-939B-41	Sequence 41, Appl
41	15	46.9	6	4	US-08-464-000-41	Sequence 41, Appl
42	15	46.9	6	4	US-09-187-859-650	Sequence 650, App
43	15	46.9	6	5	PCT-US95-10793-4	Sequence 4, Appl
44	15	46.9	7	1	US-08-136-743B-6	Sequence 6, Appl
45	15	46.9	7	1	US-08-136-743B-26	Sequence 26, Appl

#### ALIGNMENTS

RESULT 1  
US-09-187-859-637  
Sequence 637, Application US/09187859A  
Patent No. 6358920  
GENERAL INFORMATION:  
APPLICANT: Blaschuk, Orest W.  
APPLICANT: Gou, Barbara J.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL  
TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS  
FILE REFERENCE: 100086.407C1  
CURRENT APPLICATION NUMBER: US/09/187.859A  
CURRENT FILING DATE: 1998-11-06  
NUMBER OF SEQ ID NOS: 4052  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 637  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Representative linear modulating agent based on  
OTHER INFORMATION: cadherin-related neuronal receptor cell adhesion  
US-09-187-859-637

Query Match 56.2%; Score 18; DB 4; Length 7;  
Best Local Similarity 57.1%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FTITISS 7  
Db 1 FTITISS 7

RESULT 2  
US-08-757-177-16  
Sequence 16, Application US/08757177  
Patent No. 6071718  
GENERAL INFORMATION:  
APPLICANT: MUKERJI, PRADEEP  
APPLICANT: HARDS, ROBERT G.  
APPLICANT: THURMOND, JENNIFER M.  
TITLE OF INVENTION: LEONARD, AMANDA EUN-YEONG  
METHODS OF PRODUCING A RECOMBINANT PROTEIN  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESSES:  
ADDRESSSEE: Abbott Laboratories  
STREET: One Abbott Park Road  
CITY: Abbott Park  
STATE: Illinois

COUNTRY: USA  
ZIP: 60064  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/757,177  
FILING DATE: 27-NOV-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6004.US.01  
TELEPHONE: 847-935-1729  
TELEFAX: 848-938-2623  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-757-177-16

Query Match 53.18; Score 17; DB 3; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 TISS 7  
Db 2 TISS 5

RESULT 3  
US-08-739-401A-6  
Sequence 6, Application US/08739401A  
Patent No. 5837461  
GENERAL INFORMATION:  
APPLICANT: Neltz, Maureen E.  
TITLE OF INVENTION: DETECTION OF CONE-PHOTORECEPTOR-BASED  
TITLE OF INVENTION: VISION DISORDERS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Charles & Brady  
STREET: 411 East Wisconsin Avenue  
CITY: Milwaukee  
STATE: Wisconsin  
COUNTRY: U.S.A.  
ZIP: 53202-4497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/739,401A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Baker, Jean C.  
REGISTRATION NUMBER: 35,433  
REFERENCE/DOCKET NUMBER: 650053.91151  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 271-3552  
TELEFAX: (414) 271-3552  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-739-401A-6

Query Match 53.18; Score 17; DB 2; Length 7;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTTT 4  
Db 2 FTTT 5

RESULT 4  
US-08-860-904-9  
Sequence 9, Application US/08860904  
Patent No. 6294654  
GENERAL INFORMATION:  
APPLICANT: Sandlie, Inger  
APPLICANT: Bogen, Bjørne  
APPLICANT: Fossum, Sigbjørn  
TITLE OF INVENTION: A Modified Immunoglobulin Molecule  
TITLE OF INVENTION: Incorporating an Antigen in a NO. 6294654-CDR  
TITLE OF INVENTION: Loop Region  
FILE REFERENCE: 9914-1  
CURRENT APPLICATION NUMBER: US/08/860,904  
CURRENT FILING DATE: 1997-09-29  
EARLIER APPLICATION NUMBER: PCT/GB96/00116  
EARLIER FILING DATE: 1996-01-19  
EARLIER APPLICATION NUMBER: GB 9501079.9  
EARLIER FILING DATE: 1995-01-19  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 9  
LENGTH: 4  
TYPE: PRT  
ORGANISM: Mus sp.  
US-08-860-904-9

Query Match 50.0%; Score 16; DB 4; Length 4;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 TISS 7  
Db 1 TISS 4

RESULT 5  
US-08-252-995D-7  
Sequence 7, Application US/08252995D  
Patent No. 5650501  
GENERAL INFORMATION:  
APPLICANT: Denoils, James W  
APPLICANT: Heferman, Mike  
APPLICANT: Rode, Carol  
TITLE OF INVENTION: NOVEL SERINE/THREONINE KINASE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERESKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS



SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/252,995D  
FILING DATE: 02-JUN-1994  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurdydk, Linda M  
REGISTRATION NUMBER: 34,971  
REFERENCE/DOCKET NUMBER: 3153-96  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Mus musculus  
US-08-252-995D-7

Query Match 50.0%; Score 16; DB 1; Length 6;  
Best Local Similarity 60.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LTIS 7  
11:1  
Db 2 LTLSN 6

RESULT 6  
US-08-478-386A-46  
Sequence 46, Application US/08478386A  
Patent No. 5830462  
GENERAL INFORMATION:  
APPLICANT: Crabtree, Gerald R.  
APPLICANT: Schreiber, Stuart L.  
APPLICANT: Wandless, Thomas J.  
APPLICANT: Belshaw, Peter  
TITLE OF INVENTION: REGULATED TRANSCRIPTION OF TARGETED  
TITLE OF INVENTION: GENES AND OTHER BIOLOGICAL EVENTS  
NUMBER OF SEQUENCES: 81  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARIAD Pharmaceuticals, Inc.  
STREET: 26 Landsdowne Street  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02139  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC/DOS/MS/DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,386A  
FILING DATE: 07/JUN/1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Figg, E. Anthony  
REGISTRATION NUMBER: 27,195  
REFERENCE/DOCKET NUMBER: 2054-114A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 783-6040  
TELEFAX: (202) 783-6031  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-478-386A-46

Query Match 50.0%; Score 16; DB 2; Length 6;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LTIS 6  
11:1  
Db 3 LTVS 6

RESULT 7  
US-08-292-597-46  
Sequence 46, Application US/08292597  
Patent No. 5834266  
GENERAL INFORMATION:  
APPLICANT: Gerald R. Crabtree  
APPLICANT: Schreiber, Stuart L.  
APPLICANT: Wandless, Thomas J.  
APPLICANT: Belshaw, Peter  
TITLE OF INVENTION: Regulated Apoptosis  
NUMBER OF SEQUENCES: 81  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARIAD Pharmaceuticals, Inc.  
STREET: 26 Landsdowne Street  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02139  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC/DOS/MS/DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/292,597  
FILING DATE: 18/AUG/1994  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Figg, E. Anthony  
REGISTRATION NUMBER: 27,195  
REFERENCE/DOCKET NUMBER: 2054-108A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 783-6040  
TELEFAX: (202) 783-6031  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-292-597-46

Query Match 50.0%; Score 16; DB 2; Length 6;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LTIS 6  
11:1  
Db 3 LTVS 6

RESULT 8  
US-08-388-653-46  
Sequence 46, Application US/08388653

Patent No. 5869337  
GENERAL INFORMATION:  
APPLICANT: Crabtree, Gerald R.  
APPLICANT: Schreiber, Stuart L.  
APPLICANT: Spencer, David M.  
APPLICANT: Wendless, Thomas J.  
APPLICANT: Belshaw, Peter  
TITLE OF INVENTION: REGULATED TRANSCRIPTION OF TARGETED  
NUMBER OF SEQUENCES: 81  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARIAD Pharmaceuticals, Inc.  
STREET: 26 landsdowne Street  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02139  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC/DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/388,653  
FILING DATE: 14-FEB-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/478,386  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Fligg, E. Anthony  
REGISTRATION NUMBER: 27,195  
REFERENCE/DOCKET NUMBER: 2054-114A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 783-6040  
TELEFAX: (202) 783-6040  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-388-653-46

Query Match 50.0%; Score 16; DB 2; Length 6;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LTIS 6  
||:|  
Db 3 LTVS 6

RESULT 9  
US-08-473-985-46  
Sequence 46, Application US/08473985  
Patent No. 5871753  
GENERAL INFORMATION:  
APPLICANT: Crabtree, Gerald R.  
APPLICANT: Schreiber, Stuart L.  
APPLICANT: Spencer, David M.  
APPLICANT: Wendless, Thomas J.  
APPLICANT: Belshaw, Peter  
APPLICANT: Ho, Steffan  
TITLE OF INVENTION: Regulated Transcription of Targeted Genes and  
TITLE OF INVENTION: Other Biological Events  
NUMBER OF SEQUENCES: 66  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California

COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473,985  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/179,748  
FILING DATE: 07-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-5U 9863  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-473-985-46

Query Match 50.0%; Score 16; DB 2; Length 6;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LTIS 6  
||:|  
Db 3 LTVS 6

RESULT 10  
US-08-834-108-7  
Sequence 7, Application US/08834108  
Patent No. 5976893  
GENERAL INFORMATION:  
APPLICANT: Dennis, James W  
APPLICANT: Heffernan, Mike  
APPLICANT: Fode, Carol  
TITLE OF INVENTION: NOVEL SERINE/THREONINE KINASE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERESKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/834,108  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurdzyk, Linda M  
REGISTRATION NUMBER: 34,971  
REFERENCE/DOCKET NUMBER: 3153-210  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7811  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Mus musculus  
US-08-834-108-7

Query Match 50.0%; Score 16; DB 2; Length 6;  
Best Local Similarity 60.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 LTIS 7  
11:1  
DB 2 LTISN 6

RESULT 11  
US-08-483-898-46  
; Sequence 46, Application US/08483898  
; Patent No. 5994313  
; GENERAL INFORMATION:  
; APPLICANT: Gerald R. Crabtree  
; APPLICANT: Schreiber, Stuart L.  
; APPLICANT: Spencer, David M.  
; APPLICANT: Wandless, Thomas J.  
; APPLICANT: Belshaw, Peter  
; TITLE OF INVENTION: Regulated Apoptosis  
; NUMBER OF SEQUENCES: 81  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ARIAD Pharmaceuticals, Inc.  
; STREET: 26 Landsdowne Street  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02139  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/483, 898  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/292, 597  
; FILING DATE: 18-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fligg, E. Anthony  
; REGISTRATION NUMBER: 27, 195  
; REFERENCE/DOCKET NUMBER: 2054-108A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 783-6040  
; TELEFAX: (202) 783-6031  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-483-898-46

Query Match 50.0%; Score 16; DB 2; Length 6;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 LTIS 6  
11:1

DB 3 LTIS 6

RESULT 12  
US-09-087-716-46  
; Sequence 46, Application US/09087716  
; Patent No. 6011018  
; GENERAL INFORMATION:  
; APPLICANT: Crabtree, Gerald R.  
; APPLICANT: Schreiber, Stuart L.  
; APPLICANT: Spencer, David M.  
; APPLICANT: Wandless, Thomas J.  
; APPLICANT: Belshaw, Peter  
; TITLE OF INVENTION: REGULATED TRANSCRIPTION OF TARGETED  
; GENES AND OTHER BIOLOGICAL EVENTS  
; NUMBER OF SEQUENCES: 81  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ARIAD Pharmaceuticals, Inc.  
; STREET: 26 Landsdowne Street  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02139  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/087, 716  
; FILING DATE:  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 08/388, 653  
; FILING DATE: 02/14/1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fligg, E. Anthony  
; REGISTRATION NUMBER: 27, 195  
; REFERENCE/DOCKET NUMBER: 2054-114A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 783-6040  
; TELEFAX: (202) 783-6031  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-087-716-46

Query Match 50.0%; Score 16; DB 3; Length 6;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 LTIS 6  
11:1  
DB 3 LTIS 6

RESULT 13  
US-09-157-753-46  
; Sequence 46, Application US/09157753  
; Patent No. 6043082  
; GENERAL INFORMATION:  
; APPLICANT: Crabtree, Gerald R.  
; APPLICANT: Schreiber, Stuart L.  
; APPLICANT: Spencer, David M.  
; APPLICANT: Wandless, Thomas J.  
; APPLICANT: Belshaw, Peter  
; TITLE OF INVENTION: REGULATED TRANSCRIPTION OF TARGETED  
; GENES AND OTHER BIOLOGICAL EVENTS  
; NUMBER OF SEQUENCES: 81

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: ARIAD Pharmaceuticals, Inc.  
;; STREET: 26 Landsdowne Street  
;; CITY: Cambridge  
;; STATE: Massachusetts  
;; COUNTRY: USA  
;; ZIP: 02139  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC/DOS/MS/DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/157,753  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/388,653  
;; FILING DATE: 14-FEB-1995  
;; APPLICATION NUMBER: US 08/478,386  
;; FILING DATE: 07-JUN-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Flagg, E. Anthony  
;; REGISTRATION NUMBER: 27,195  
;; REFERENCE/DOCKET NUMBER: 2054-114A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 783-6040  
;; TELEFAX: (202) 783-6031  
;; INFORMATION FOR SEQ ID NO: 46:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 6 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-09-157-753-46

Query Match 50.0%; Score 16; DB 3; Length 6;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LITS 6  
||:|  
Db 3 LITS 6

RESULT 14  
US-09-157-230-46  
;; Sequence 46, Application US/09157230  
;; Patent No. 6046047  
;; GENERAL INFORMATION:  
;; APPLICANT: Crabtree, Gerald R.  
;; APPLICANT: Schreiber, Stuart L.  
;; APPLICANT: Spencer, David M.  
;; APPLICANT: Wandless, Thomas J.  
;; APPLICANT: Belshaw, Peter  
;; TITLE OF INVENTION: REGULATED TRANSCRIPTION OF TARGETED  
;; TITLE OF INVENTION: GENES AND OTHER BIOLOGICAL EVENTS  
;; NUMBER OF SEQUENCES: 81  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: ARIAD Pharmaceuticals, Inc.  
;; STREET: 26 Landsdowne Street  
;; CITY: Cambridge  
;; STATE: Massachusetts  
;; COUNTRY: USA  
;; ZIP: 02139  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC/DOS/MS/DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/157,230

;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/478,386  
;; FILING DATE: 07/JUN/1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Flagg, E. Anthony  
;; REGISTRATION NUMBER: 27,195  
;; REFERENCE/DOCKET NUMBER: 2054-114A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 783-6040  
;; TELEFAX: (202) 783-6031  
;; INFORMATION FOR SEQ ID NO: 46:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 6 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-09-157-230-46

Query Match 50.0%; Score 16; DB 3; Length 6;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LITS 6  
||:|  
Db 3 LITS 6

RESULT 15  
US-09-087-811-46  
;; Sequence 46, Application US/09087811  
;; Patent No. 6054436  
;; GENERAL INFORMATION:  
;; APPLICANT: Gerald R. Crabtree  
;; APPLICANT: Schreiber, Stuart L.  
;; APPLICANT: Spencer, David M.  
;; APPLICANT: Wandless, Thomas J.  
;; APPLICANT: Belshaw, Peter  
;; TITLE OF INVENTION: Regulated Apoptosis  
;; NUMBER OF SEQUENCES: 81  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: ARIAD Pharmaceuticals, Inc.  
;; STREET: 26 Landsdowne Street  
;; CITY: Cambridge  
;; STATE: Massachusetts  
;; COUNTRY: USA  
;; ZIP: 02139  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC/DOS/MS/DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/087,811  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/292,597  
;; FILING DATE: 18-AUG-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Flagg, E. Anthony  
;; REGISTRATION NUMBER: 27,195  
;; REFERENCE/DOCKET NUMBER: 2054-108A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202) 783-6040  
;; TELEFAX: (202) 783-6031  
;; INFORMATION FOR SEQ ID NO: 46:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 6 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear

; MOLECULE TYPE: protein  
US-09-087-811-46

Query Match 50.0%; Score 16; DB 3; Length 6;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0;  
Gaps 0;

QY 3 LTIS 6  
||:|  
Db 3 LTVS 6

Search completed: July 15, 2002, 13:26:12  
Job time: 492 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:27:07 ; Search time 49.13 Seconds

(without alignments)  
13.691 Million cell updates/sec

Title: US-09-712-819A-2

Perfect score: 33  
Sequence: 1 TRIFSXL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 28138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 455

Minimum DB seq length: 0  
Maximum DB seq length: 7

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12	36.4	6	2 A61411	ameletin - rat
2	12	36.4	6	2 I59142	platelet-derived g
3	12	36.4	7	2 A30812	sex pheromone cCFL
4	10	30.3	4	2 I40505	hypothetical prote
5	10	30.3	5	2 PT0295	Ig heavy chain CRD
6	10	30.3	6	2 A61068	locustaxidin - mig
7	10	30.3	6	2 I49421	lamnin 81 - weste
8	10	30.3	6	2 PD0028	pev-kinin 2 - pena
9	10	30.3	7	2 JN0859	peptidyl-di-peptida
10	10	30.3	7	2 A25269	sex pheromone CAM3
11	10	30.3	7	2 PX0008	glucuronosyltransf
12	10	30.3	7	2 PD0029	pev-kinin 1 - pena
13	10	30.3	7	2 I55382	hypothetical pepti
14	9	27.3	4	2 T46627	hypothetical prote
15	9	27.3	4	2 J01273	neuropetide Antho
16	9	27.3	4	2 E44823	synaptosomal-assoc
17	9	27.3	4	2 PT0534	T-cell receptor be
18	9	27.3	5	2 C41225	cooper resistance
19	9	27.3	5	2 I40702	primase - Citrodac
20	9	27.3	5	2 S11127	phosphoprotein, bo
21	9	27.3	5	2 S62883	seminal plasma, pro
22	9	27.3	6	2 A60986	N-formyl oligopept
23	9	27.3	6	2 A60494	antineoplastic gly
24	9	27.3	6	2 I37263	Y protein - human
25	9	27.3	6	2 I48126	alpha-tubulin - Ch
26	9	27.3	6	2 B33932	Ig mu chain D regi
27	9	27.3	6	2 I49424	cytotoxic T-lympho
28	9	27.3	7	2 B39127	phosphotransferase
29	9	27.3	7	2 I40504	hypothetical prote

30	9	27.3	7	2 S68004	hucolin, 75K chain
31	8	24.2	4	2 PT0645	T-cell receptor be
32	8	24.2	5	2 A44692	fulicin - giant Af
33	8	24.2	5	2 PT0278	Ig heavy chain CRD
34	8	24.2	5	2 PT0624	T-cell receptor be
35	8	24.2	5	2 PT0525	T-cell receptor be
36	8	24.2	5	2 PT0713	T-cell receptor be
37	8	24.2	6	2 I51317	hILH transcription
38	8	24.2	6	2 PT0519	T-cell receptor be
39	8	24.2	6	2 PT0621	T-cell receptor be
40	8	24.2	6	2 PT0730	T-cell receptor be
41	8	24.2	7	2 PH1408	Ig heavy chain V r
42	8	24.2	7	2 S29735	polyphosphate--glu
43	8	24.2	7	2 A34818	vicilin 72K chain
44	8	24.2	7	2 PN0150	omega-glutamine 1'
45	8	24.2	7	2 S33567	tubulin beta-3 cha

## ALIGNMENTS

RESULT 1  
A61411  
ameletin - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 09-Sep-1994 #sequence\_revision 09-Sep-1994 #text\_change 18-Aug-2000  
C:Accession: A61411  
R:Burzynski, S.R.  
Anal. Biochem. 70, 359-365, 1976  
A:Title: Sequential analysis in subnanomolar amounts of peptides. Determination of th  
A:Reference number: A61411; MUID:76182447  
A:Accession: A61411  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-6 <BUP>  
C:Superfamily: unassigned animal peptides  
C:Keywords: pyrrolidone carboxylic acid (Cln) #status experimental  
E:/Modified site: pyrrolidone carboxylic acid (Cln) #status experimental

Query Match 36.4% Score 12; DB 2; Length 6;  
Best Local Similarity 66.7% Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 FSK 6  
DB 4 YSK 6

RESULT 2  
I59142  
platelet-derived growth factor B chain - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
C:Accession: I59142  
R:Peck, M.; Gazit, A.; Arnstein, P.; Aaronson, S.A.  
Proc. Natl. Acad. Sci. U.S.A. 86, 2693-2697, 1989  
A:Title: Generation of fibrosarcomas in vivo by a retrovirus that expressed the norma  
A:Reference number: I59142; MUID:89202393  
A:Accession: I59142  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-6 <RES>  
A:Cross-references: GB:M26180; NID:g516624; PIDN:AAA39905.1; PID:g516625

Query Match 36.4% Score 12; DB 2; Length 6;  
Best Local Similarity 40.0% Pred. No. 2.8e+05;  
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 IFSKL 7  
DB 2 VFAVL 6

RESULT 3  
A30812  
sex pheromone ccf10 - *Enterococcus faecalis*  
C:Species: *Enterococcus faecalis*  
C>Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 18-Jun-1993  
C:Accession: A30812  
R:Mori, M.; Sakagami, Y.; Ishii, Y.; Isogai, A.; Kikade, C.; Fujino, M.; Adachi, J.C.; Du  
J. Biol. Chem. 263:14574-14576, 1988  
A:Title: Structure of ccf10, a peptide sex pheromone which induces conjugative transfer  
A:Reference number: A30812; MUID:93008313  
A:Accession: A30812  
A:Molecule type: Protein  
A:Residues: 1-7 <MOR>

Query Match 36.4%; Score 12; DB 2; Length 7;  
Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TRIP 4  
DB 3 TLVF 6

RESULT 4  
I40505  
hypothetical protein 3 (4 aa) - *Bacillus stearothermophilus*  
C:Species: *Bacillus stearothermophilus*  
C>Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 15-Oct-1999  
C:Accession: I40505  
R:Way, M.M.; Winter, G.  
Eur. J. Biochem. 158, 505-510, 1986  
A:Title: A transcription terminator in the 5' non-coding region of the tyrosyl tRNA synth  
A:Reference number: I40503; MUID:86274732  
A:Accession: I40505  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-4 <RES>  
A:Cross-references: EMBL:X04193; NID:q40233; PIDN:CAA27763.1; PID:q580944

Query Match 30.3%; Score 10; DB 2; Length 4;  
Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 IFSK 6  
DB 1 MLSK 4

RESULT 5  
PT0295  
Ig heavy chain CRO3 region (clone 5-91) - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Aug-1996  
C:Accession: PT0295  
R:Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.  
J. Exp. Med. 173, 395-407, 1991  
A:Title: Preferential utilization of specific immunoglobulin heavy chain diversity and  
A:Reference number: PT0222; MUID:91108317  
A:Accession: PT0295  
A:Molecule type: DNA  
A:Residues: 1-5 <YAM>  
A:Experimental source: B lymphocyte  
C:Keywords: heterotrimer; immunoglobulin

Query Match 30.3%; Score 10; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TR 2  
DB 3 TR 4

RESULT 6  
A61068  
locustakinin - migratory locust  
C:Species: *Locusta migratoria* (migratory locust)  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 08-Dec-1995  
C:Accession: A61068  
R:Schroofs, L.; Holman, G.M.; Proost, P.; Van Damme, J.; Hayes, T.K.; De Loof, A.  
Regul. Pept. 37, 43-57, 1992  
A:Title: Locustakinin, a novel myotropic peptide from *Locusta migratoria*, isolation.  
A:Reference number: A61068; MUID:92262851  
A:Accession: A61068  
A:Molecule type: protein  
A:Residues: 1-6 <SCH>  
C:Keywords: amidated carboxyl end; cephalomyotropic peptide; neuropeptide  
P:6/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 30.3%; Score 10; DB 2; Length 6;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FS 5  
DB 2 FS 3

RESULT 7  
I49421  
laminin B1 - western wild mouse (fragment)  
C:Species: *Mus spretus* (western wild mouse)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
C:Accession: I49421  
R:Ko, M.S.; Wang, X.; Horton, J.H.; Hagen, M.D.; Takahashi, N.; Maezaki, Y.; Nadeau,  
Mamm. Genome 5, 349-355, 1994  
A:Title: Genetic mapping of 40 cDNA clones on the mouse genome by PCR.  
A:Reference number: I48934; MUID:94319082  
A:Accession: I49421  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-6 <RES>  
A:Cross-references: EMBL:U05736; NID:q497073; PIDN:AAB60477.1; PID:q642829

Query Match 30.3%; Score 10; DB 2; Length 6;  
Best Local Similarity 33.3%; Pred. No. 2.8e+05;  
Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 IFS 5  
DB 1 VYS 3

RESULT 8  
PD0028  
pev-kinin 2 - penaeid shrimp (*Penaeus vannamei*) (fragment)  
C:Species: *Penaeus vannamei*  
C>Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 19-May-2000  
C:Accession: PD0028  
R:Nieto, J.; Veeleert, D.; Derau, R.; Waelens, E.; Cersiaens, A.; Coast, G.; Devree  
Biochem. Biophys. Res. Commun. 248, 405-411, 1998  
A:Title: Identification of one tachykinin- and two kinin-related peptides in the brain  
A:Reference number: PD0027; MUID:98342103  
A:Accession: PD0028  
A:Molecule type: protein  
A:Residues: 1-6 <NIE>  
C:Comment: This peptide belongs to myotropic neuropeptides.



Query Match 30.3%; Score 10; DB 2; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FS 5  
 11  
 DB 2 FS 3

RESULT 9

JN0859  
 pepidyl-di-peptidase A inhibitory peptide C105 - striped bonito  
 C:Species: Sarda orientalis (striped bonito)  
 C:Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 07-May-1999  
 C:Accession: JN0859  
 R:Matsumura, N.; Fujii, M.; Takeda, Y.; Shimizu, T.  
 Biosci. Biotechnol. Biochem. 57, 1743-1744, 1993  
 A:Title: Isolation and characterization of angiotensin I-converting enzyme inhibitory pe  
 A:Reference number: JN0859; MUID:94080036  
 A:Accession: JN0859  
 A:Molecule type: protein  
 A:Residues: 1-7 <MAP>  
 A:Experimental source: Intestine  
 C:Comment: The carboxyl-terminus is essential for the protein's expression of angiotensin  
 C:Superfamily: bradykinin-potentiating peptide  
 C:Keywords: angiotensin-converting enzyme inhibitor

Query Match 30.3%; Score 10; DB 2; Length 7;  
 Best Local Similarity 66.7%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 5 SKL 7  
 11  
 DB 3 AKL 5

RESULT 10

A25269  
 sex pheromone CAM373 - Enterococcus faecalis  
 N:Alternate names: clumping-inducing agent (CIA)  
 C:Species: Enterococcus faecalis  
 C:Date: 31-Mar-1998 #sequence\_revision 31-Mar-1998 #text\_change 18-Jun-1993  
 C:Accession: A25269  
 R:Morit, M.; Tanaka, H.; Sakagami, Y.; Isogai, A.; Fujino, M.; Kitada, C.; White, B.A.; A  
 FEBS Lett. 206, 69-72, 1986  
 A:Title: Isolation and structure of the streptococcus faecalis sex pheromone, CAM373.  
 A:Reference number: A25269; MUID:87005252  
 A:Accession: A25269  
 A:Molecule type: protein  
 A:Residues: 1-7 <MOR>

Query Match 30.3%; Score 10; DB 2; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 IF 4  
 11  
 DB 2 IF 3

RESULT 11

PX0008  
 glucuronosyltransferase (EC 2.4.1.17), hepatic - rat (fragment)  
 N:Alternate names: UDP-glucuronosyltransferase  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 07-Feb-1997  
 C:Accession: PX0008  
 R:Yokota, H.; Yuasa, A.; Salo, R.  
 J. Biochem. 104, 531-536, 1988  
 A:Title: Purification and properties of a form of UDP-glucuronosyltransferase from liver m  
 A:Reference number: PX0008; MUID:89197852

A:Accession: PX0008  
 A:Molecule type: protein  
 A:Residues: 1-7 <YOK>  
 C:Keywords: glycosyltransferase; hexosyltransferase; liver

Query Match 30.3%; Score 10; DB 2; Length 7;  
 Best Local Similarity 66.7%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 5 SKL 7  
 11  
 DB 1 TKL 3

RESULT 12

PD0029  
 pty-kinin I - penaeid shrimp (Penaeus vannamei) (fragment)  
 C:Species: Penaeus vannamei  
 C:Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 19-May-2000  
 C:Accession: PD0029  
 R:Nieto, J.; Veelaert, D.; Derua, R.; Maelkens, E.; Cersiliaens, A.; Coast, G.; Devree  
 Blochem. Biophys. Res. Commun. 248, 406-411, 1998  
 A:Title: Identification of one tachykinin- and two kinin-related peptides in the brai  
 A:Reference number: PD0029; MUID:98342103  
 A:Accession: PD0029  
 A:Molecule type: protein  
 A:Residues: 1-7 <NIE>  
 C:Comment: This peptide belongs to myotropic neuropeptides.

Query Match 30.3%; Score 10; DB 2; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FS 5  
 11  
 DB 3 FS 4

RESULT 13

I53382  
 hypothetical peptide PAII promoter region - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 16-Apr-1999 #sequence\_revision 16-Apr-1999 #text\_change 20-Apr-2000  
 C:Accession: I53382  
 R:Dawson, S.J.; Wilman, B.; Hamslen, A.; Green, F.; Humphries, S.; Henney, A.M.  
 J. Biol. Chem. 268, 10739-10745, 1993  
 A:Title: The two allele sequences of a common polymorphism in the promoter of the pla  
 A:Reference number: I53382; MUID:93265509  
 A:Accession: I53382  
 A>Status: translation not shown; translated from GB/EML/DBDU  
 A:Molecule type: DNA  
 A:Residues: 1-7 <DAN>  
 A:Cross-references: GB:M91557; NID:q190020; PIRN:AAA60110.1; PID:q190021  
 C:Comment: This is the hypothetical translation of a sequence from the PAII gene prom  
 A:Cross-references: GDB:120297; OMIM:173360  
 A:Map position: 7q21.3-7q22

Query Match 30.3%; Score 10; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TR 2  
 11  
 DB 2 TR 3

RESULT 14

T46627

hypothetical protein c4 - loblolly pine  
C:Species: Pinus taeda (loblolly pine)  
C:Date: 18-Feb-2000 #sequence\_revision 18-Feb-2000 #text\_change 18-Feb-2000  
C:Accession: T46627  
R:Chang, S.; Puryear, J.; Funkhouser, E.A.; Newton, R.J.; Cairney, J.  
A:Description: Cloning of a chitinase homolog which lacks chitin binding sites and is dc  
A:Reference number: 233105  
A:Accession: T46627  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-4 <CH>  
A:Cross-references: EMBL:U31309; NID:g974285; PID:g974292  
A:Experimental source: strain 56PT2x56PT3; 8 month seedlings

Query Match 27.3% Score 9; DB 2; Length 4;  
Best local similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 6 KL 7  
11  
DB 2 KL 3

RESULT 15  
J01273  
neuropeptide Antho-Kamide - sea anemone (Anthopleura elegantissima)  
C:Species: Anthopleura elegantissima  
C:Date: 31-Mar-1992 #sequence\_revision 04-Dec-1992 #text\_change 08-Dec-1995  
C:Accession: J01273  
R:Nothacker, H.P.; Rinehart, K.L.; Grimmelikhuijsen, C.J.P.  
Biochem. Biophys. Res. Commun. 179, 1205-1211, 1991  
A:Title: Isolation of L-3-phenyllactyl-Phe-Lys-Ala-NH2 (Antho-Kamide), a novel neuropep  
A:Reference number: J01273; MUID:92028952  
A:Accession: J01273  
A:Molecule type: protein  
A:Residues: 1-4 <NOT>  
C:Comment: The carboxyl-terminal amide probably arises from cleavage of a following glyc  
C:Keywords: amidated carboxyl end; neuropeptide; phenyllactylation  
F:1/Modified site: L-3-phenyllactic acid (Phe) #status experimental  
F:4/Modified site: amidated carboxyl end (Ala) #status experimental

Query Match 27.3% Score 9; DB 2; Length 4;  
Best local similarity 66.7%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 4 FSK 6  
11  
DB 1 FFK 3

Search completed: July 15, 2002, 13:27:08  
Job time: 438 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:38:10 ; Search time 24.49 Seconds  
(without alignments)

11.067 Million cell updates/sec

Title: US-09-712-819A-2

Perfect score: 33

Sequence: 1 TRIFSKL 7

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 84

Minimum DB seq length: 0  
Maximum DB seq length: 7

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12	36.4	7	1	CCP1_ENTFA
2	10	30.3	5	1	UC22_MAZE
3	10	30.3	6	1	LOK1_LOCM1
4	10	30.3	7	1	CIA_ENTFA
5	9	27.3	4	1	FFKA_ANTEL
6	9	27.3	4	1	FYRL_ANTEL
7	9	27.3	5	1	PSK_DAUCA
8	9	27.3	7	1	UN06_PINPS
9	8	24.2	7	1	PARR_PROCL
10	8	24.2	7	1	PARR_PROCL
11	7	21.2	4	1	ACH1_ACHRU
12	7	21.2	4	1	EMR_MACNT
13	7	21.2	4	1	TUPT_HUMAN
14	7	21.2	5	1	AL14_CARMA
15	7	21.2	5	1	RE11_LITRU
16	7	21.2	5	1	RE21_LITRU
17	7	21.2	5	1	RE31_LITRU
18	7	21.2	5	1	RE32_LITRU
19	7	21.2	6	1	CIP1_MYTED
20	7	21.2	6	1	CIP1_MYTED
21	7	21.2	6	1	TM0F_SARBU
22	7	21.2	7	1	ALU7_CTDPO
23	7	21.2	7	1	CAR_MYTED
24	7	21.2	7	1	E105_LITRU
25	7	21.2	7	1	FARB_CALVO
26	7	21.2	7	1	HY7_PIG
27	7	21.2	7	1	LANC_CARUT
28	7	21.2	7	1	TY51_LITRU
29	6	18.2	4	1	DCMS_PSECH
30	6	18.2	4	1	FAR3_HIRME
31	6	18.2	4	1	FAR4_HIRME
32	6	18.2	4	1	FLRN_HIRME
33	6	18.2	4	1	FLRN_ANTEL

34	6	18.2	4	1	OCPI_OCTMI	P58648 octopus min
35	6	18.2	5	1	E103_LITRU	P82099 litorea rub
36	6	18.2	5	1	E104_LITRU	P82100 litorea rub
37	6	18.2	5	1	FARP_ARTTR	P41853 artroposthi
38	6	18.2	5	1	PAP2_PAPMA	P41854 artroposthi
39	6	18.2	5	1	SUGA_ACHDO	P19991 acheta dome
40	6	18.2	5	1	TPIS_CANFA	P54714 canis famli
41	6	18.2	6	1	E101_LITRU	P82096 litorea rub
42	6	18.2	6	1	FARP_MONEX	P41966 monilezia ex
43	6	18.2	7	1	ALL2_CARMA	P81805 carcinus ma
44	6	18.2	7	1	ALL3_CARMA	P81806 carcinus ma
45	6	18.2	7	1	ALL4_CARMA	P81807 carcinus ma

#### ALIGNMENTS

RESULT 1  
ID CCP1\_ENTFA STANDARD: PRT: 7 AA.  
AC P20104;  
DF 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-FEB-1991 (Rel. 17, Last annotation update)  
DE Sex pheromone CCP10. (Streptococcus faecalis).  
OS Enterococcus faecalis  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;  
OX NCBI\_TaxID=1351;  
RN 11  
RP SEQUENCE.  
RX MEDLINE=89008313; PubMed=3139658;  
RA Moril M., Sakagami Y., Ishii Y., Isogai A., Kitada C., Fujino M.,  
RA Adail J.C., Danny G.M., Suzuki A.;  
RT "Structure of CCP10, a peptide sex pheromone which induces  
RT conjugative transfer of the Streptococcus faecalis tetracycline  
RT resistance plasmid, pCCF10.";  
RL J. Biol. Chem. 263:14574-14578(1988).  
CC -1- FUNCTION: CCP10 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE  
CC PIR: A30812; A30812.  
DR HEMOLISIN PLASMID PCF10.  
SQ PHEROMONE.  
KW SEQUENCE 7 AA: 790 MW: 72C9DC731B2C740 CRC64:

Query Match 36.4%; Score 12; DB 1; Length 7;  
Best Local Similarity 50.0%; Pred. No. 1e+05;  
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TRIF 4  
DB 3 TLVP 6  
RESULT 2  
ID UC22\_MAZE STANDARD: PRT: 5 AA.  
AC P80628;  
DF 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Unknown protein from 2D-page of elutriated coleoptile (Spot 474)  
DE (Fragment).  
OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Panicoideae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN 11  
RP SEQUENCE.  
RC TISSUE-Coleoptile;  
RC Touzet P., Riccardi F., Morin C., Damerval C., Huet J.-C.,  
RA Permollet J.-C., Zivy M., de Vienne D.;

RT "The maize two dimensional gel protein database: towards an integrated  
genome analysis program."  
RL Theor. Appl. Genet. 93:997-1005(1996).  
CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN  
PROTEIN IS: 6.1, ITS MW IS: 30.4 kDa.  
DR Maize-2DPAGE: P80628; COLEOPTILE.  
DR MaizeDB: 123954; -.  
FT NON\_TER 1  
FT MOD\_RES 5  
SQ SEQUENCE 5 AA: 654 MW: 72CB19C9C0300000 CRC64;

Query Match 30.3%; Score 10; DB 1; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 IF 4  
DB 1 IF 2

RESULT 3  
ID LOKI.LOCMT STANDARD; PRT: 6 AA.  
AC P41491;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Locustakinin I.  
OS Locusta migratoria (Migratory locust).  
OC Eukaryota; Metazoa; Arthropoda; Insecta;  
OC Pterygota; Neoptera; Orthoptera; Caelifera;  
OC Acridoidea; Acrididae; Locusta.  
OC NCBI\_TaxID=7004;  
RN 11  
RP SEQUENCE.  
RC TISSUE=Corpora cardiaca;  
RX MEDLINE=92262851; Pubmed=1585017;  
RA Schoofs L., Holman G.M., Proost P., van Damme J., Hayes T.K.,  
de Loof A.,  
RT "Locustakinin, a novel myotropic peptide from Locusta migratoria,  
RT isolation, primary structure and synthesis."  
RL Regul. Pept. 37:49-57(1992).  
CC -1- FUNCTION: MYOTROPIC PEPTIDE. MAY BE IMPORTANT IN THE STIMULATION  
OF ION TRANSPORT AND INHIBITION OF DIVERGIC ACTIVITY IN MALPIGHIAN  
TUBULES.  
CC PIR: A61068; A61068.  
DR Neuropptide; Amidation.  
KW MOD\_RES 6  
SQ SEQUENCE 6 AA: 654 MW: 686365A5B9CDB000 CRC64;

Query Match 30.3%; Score 10; DB 1; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 FS 5  
DB 2 FS 3

RESULT 4  
ID CIA.ENTFA STANDARD; PRT: 7 AA.  
AC P11932;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 01-FEB-1991 (Rel. 17, Last annotation update)  
DE Sex pheromone CAM373 (clumping-inducing agent) (CIA).  
OS Enterococcus faecalis (Streptococcus faecalis).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;  
OC Enterococcus.  
OC NCBI\_TaxID=1351;

RN [1]  
RP SEQUENCE.  
RX MEDLINE=87005252; Pubmed=3093276;  
RA Mori M., Tanaka H., Sakagami Y., Isogai A., Fujino M., Kitada C.,  
RA White B.A., An F.Y., Clewell D.B., Suzuki A.;  
RT "Isolation and structure of the Streptococcus faecalis sex pheromone,  
RT CAM373."  
RL FEBS Lett. 206:69-72(1986).  
CC -1- FUNCTION: CAM373 INDUCES MATING RESPONSE OF DONOR CELLS  
CC HARBORING PAM373.  
CC -1- MISCELLANEOUS: THE N-TERMINUS IS POSSIBLY RESPONSIBLE FOR  
CC SPECIFICITY OF PHEROMONES TO PLASMIDS.  
CC -1- SIMILARITY: C-TERMINAL TO THE PHEROMONES CPD1 AND CAD1.  
DR PIR: A25269; A25269.  
KW Pheromone.  
SQ SEQUENCE 7 AA: 734 MW: 75BDD72059C05DB0 CRC64;

Query Match 30.3%; Score 10; DB 1; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 IF 4  
DB 2 IF 3

RESULT 5  
ID FPKA.ANTEL STANDARD; PRT: 4 AA.  
AC P58705;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Antho-Kamide.  
OS Anthopleura elegantissima (Sea anemone).  
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinoptaria;  
OC Nymphaeae; Actinellidae; Anthopleura.  
OC NCBI\_TaxID=6110;  
RN 11  
RP SEQUENCE.  
RX Pubmed=1681803;  
RA Nothacker H.-P., Rinehart K.L. Jr., Grimmelikhuijzen C.J.P.,  
RT "Isolation of L-3-phenyllactyl-Phe-Iys-Ala-NH2 (Antho-Kamide), a  
RT novel neuropeptide from sea anemones."  
RL Biochem. Biophys. Res. Commun. 179:1205-1211(1991).  
RN 12  
RP FUNCTION.  
RX Pubmed=8397415;  
RA McFarlane I.D., Hudman D., Nothacker H.-P., Grimmelikhuijzen C.J.P.,  
RT "The expansion behaviour of sea anemones may be coordinated by two  
RT inhibitory neuropeptides, Antho-Kamide and Antho-Ramide."  
RL Proc. R.Soc. Lond. B. Biol. Sci. 253:183-188(1993).  
CC -1- FUNCTION: Inhibits spontaneous contractions in several muscle  
CC groups. May be involved in the expansion phase of feeding  
CC behaviour in sea anemones.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- TISSUE SPECIFICITY: Neuron-specific.  
KW Neuropptide; Amidation.  
FT MOD\_RES 1  
FT MOD\_RES 4  
FT MOD\_RES 4  
SQ SEQUENCE 4 AA: 512 MW: 6DD339C9A0000000 CRC64;

Query Match 27.3%; Score 9; DB 1; Length 4;  
Best Local Similarity 66.7%; Pred. No. 1e+05;  
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 FSK 6  
DB 1 FPK 3

```

RESULT 6
FYRL ANTEL
ID FYRL ANTEL STANDARD: PRT: 4 AA.
AC P58706;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Antho-Riamide I [Contains: Antho-Riamide II]
OS Anthopleura elegantissima (Sea anemone)
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinaria;
OC Myantheae; Actiniidae; Anthopleura.
OX NCBI_TaxID=6110;
RN
RP
RX PubMed=8397415;
RA McFarlane I.D., Budman D., Nothacker H.-P., Grimelikhuijzen C.J.P.;
RT "The expansion behaviour of sea anemones may be coordinated by two
RT inhibitory neuropeptides, Antho-Kamide and Antho-Riamide.";
RL Proc. R. Soc. Lond. B, Biol. Sci. 253:183-188(1993).
CC -1- FUNCTION: Inhibits spontaneous contractions in several muscle
CC groups. May be involved in the expansion phase of feeding
CC behaviour in sea anemones.
CC -1- SUBCELLULAR LOCATION: Secreted.
KW Neuropeptide; Amidation.
FT CHAIN 1 4 ANTHO-RIAMIDE I.
FT MOD_RES 2 4 ANTHO-RIAMIDE II.
FT MOD_RES 1 1 L-3-PHENYLALANINE.
FT MOD_RES 4 4 AMIDATION.
SQ SEQUENCE 4 AA: 598 MW: 60441B59A0000000 CRC64:

Query Match 27.3%; Score 9; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RI 3
   1 1
Db 3 RI 4

RESULT 7
PSK DAUCA
ID PSK DAUCA STANDARD: PRT: 5 AA.
AC P58261;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phytosulfoxine-alpha (PSK-alpha) [Contains: Phytosulfoxine-beta (PSK-
DE beta)].
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; easterids II; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN
RP
RX PubMed=1821096;
RA Nothacker H.-P., Rinehart K.L. Jr., McFarlane I.D.,
RA Grimelikhuijzen C.J.P.;
RT "Isolation of two novel neuropeptides from sea anemones: the unusual,
RT biologically active L-3-phenylalanyl-Tyr-Arg-Ile-NH2 and its
RT des-phenylalanyl fragment Tyr-Arg-Ile-NH2.";
RL Peptides 12:1165-1173(1991).
RN
RP
RX PubMed=8397415;
RA McFarlane I.D., Budman D., Nothacker H.-P., Grimelikhuijzen C.J.P.;
RT "The expansion behaviour of sea anemones may be coordinated by two
RT inhibitory neuropeptides, Antho-Kamide and Antho-Riamide.";
RL Proc. R. Soc. Lond. B, Biol. Sci. 253:183-188(1993).
CC -1- FUNCTION: Inhibits spontaneous contractions in several muscle
CC groups. May be involved in the expansion phase of feeding
CC behaviour in sea anemones.
CC -1- SUBCELLULAR LOCATION: Secreted.
KW Neuropeptide; Amidation.
FT CHAIN 1 4 ANTHO-RIAMIDE I.
FT MOD_RES 2 4 ANTHO-RIAMIDE II.
FT MOD_RES 1 1 L-3-PHENYLALANINE.
FT MOD_RES 4 4 AMIDATION.
SQ SEQUENCE 4 AA: 598 MW: 60441B59A0000000 CRC64:

```

```

CC -1- FUNCTION: IN PRESENCE OF 2,4-D, STIMULATES PROLIFERATION OF THE
CC CELLS, BUT DOES NOT STIMULATE DIFFERENTIATION INTO THE SOMATIC
CC EMBRYOS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: SULFATION IS IMPORTANT FOR ACTIVITY AND FOR THE BINDING TO A
CC PUTATIVE MEMBRANE RECEPTOR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PHYTOSULFOXINE FAMILY.
KW Growth factor; Sulfation.
FT PEPTIDE 1 4 PHYTOSULFOXINE-BETA.
FT MOD_RES 1 1 SULFATION.
FT MOD_RES 3 3 SULFATION.
FT MOD_RES 5 AA: 687 MW: 76C1B5504B300000 CRC64:
SQ SEQUENCE 5 AA: 687 MW: 76C1B5504B300000 CRC64:

Query Match 27.3%; Score 9; DB 1; Length 5;
Best Local Similarity 25.0%; Pred. No. 1e+05;
Matches 1; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 3 IFSK 6
   1 1
Db 2 IYIQ 5

RESULT 8
UN06 PINPS
ID UN06 PINPS STANDARD: PRT: 7 AA.
AC P81675;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Unknown protein from 2D-page of needles (NL41) (Fragment).
OS Pinus pinaster (Maritime pine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=71647;
RN
RP
RX PubMed=99274088; PubMed=10344291;
RA Costa P., Plomieu C., Baw G., Dubos C., Bahman N., Kremer A.,
RA Frigerio J.-M., Plomieu C.;
RT "Separation and characterization of needle and xylem maritime pine
RT proteins.";
RL Electrophoresis 20:1098-1108(1999).
CC -1- MISCELLANEOUS: ON THE 2D-GEL, THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 6.6, ITS MW IS: 25 kDa.
FT NON_TER 1 1
FT NON_TER 7 7
FT NON_TER 7 7
SQ SEQUENCE 7 AA: 823 MW: 69D76724486B5740 CRC64:

Query Match 27.3%; Score 9; DB 1; Length 7;
Best Local Similarity 20.0%; Pred. No. 1e+05;
Matches 1; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 IFSK 7
   1 1
Db 1 LYGNL 5

RESULT 9
FYRL PROCL
ID FYRL PROCL STANDARD: PRT: 7 AA.
AC P58499;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1994 (Rel. 32, Last annotation update)
DE Cardiac excitatory FMRFamide homolog NFL.
OS Procambiarus clarkii (Red swamp crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Plesocyemata; Astacidea;
OC Astacoidae; Cambaridae; Procambiarus.
OX NCBI_TaxID=6728;

```

```

RN [1]
RP SEQUENCE:
RX TISSUE-Pericardial organs;
RX MDLINE-93248032; PubMed-8387183;
RA Mercier A.J., Orchard I., Tebrugge V., Skerrett M.;
RT "Isolation of two FMRFamide-related peptides from crayfish
   pericardial organs."
RL Peptides 14:137-143(1993).
CC -1- FUNCTION: INCREASES THE RATE AND AMPLITUDE OF SPONTANEOUS
CC CONTRACTIONS OF SEMI-ISOLATED HEARTS. INCREASES THE AMPLITUDE OF
CC EXCITATORY POSTSYNAPTIC POTENTIALS IN ABDOMINAL EXTENSOR MUSCLE.
CC -1- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC FAMILY.
KM Neuropeptide; Amidation.
FT MOD_RES 7
SQ SEQUENCE 7 AA; 966 MW; 69D40729C4540420 CRC64;

Query Match 24.2%; Score 8; DB 1; Length 7;
Best Local Similarity 66.7%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 RIF 4
DB 2 RNF 4

RESULT 10
FAR2_PROCL STANDARD; PRT; 7 AA.
AC P38498;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Cardioexcitatory FMRFamide homolog DP2.
DE Procamburus clarkii (Red swamp crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacoides; Cambaridae; Procamburus.
OX NCBI_TaxID=6728;
RN [1]
RP SEQUENCE:
RX TISSUE-Pericardial organs;
RX MDLINE-93248032; PubMed-8387183;
RA Mercier A.J., Orchard I., Tebrugge V., Skerrett M.;
RT "Isolation of two FMRFamide-related peptides from crayfish
   pericardial organs."
RL Peptides 14:137-143(1993).
CC -1- FUNCTION: INCREASES THE RATE AND AMPLITUDE OF SPONTANEOUS
CC CONTRACTIONS OF SEMI-ISOLATED HEARTS. INCREASES THE AMPLITUDE OF
CC EXCITATORY POSTSYNAPTIC POTENTIALS IN ABDOMINAL EXTENSOR MUSCLE.
CC -1- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC FAMILY.
KM Neuropeptide; Amidation.
FT MOD_RES 7
SQ SEQUENCE 7 AA; 967 MW; 69D40729C4540420 CRC64;

Query Match 24.2%; Score 8; DB 1; Length 7;
Best Local Similarity 66.7%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 RIF 4
DB 2 RNF 4

RESULT 11
ACHL_ACHFU STANDARD; PRT; 4 AA.
AC P35904;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)

```

```

DT 15-JUN-1998 (Rel. 36, Last annotation update)
DE Achatin-I.
OS Achatina fulica (giant African snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC Achatinacea; Achatinidae; Achatina.
OX NCBI_TaxID=6530;
RN [1]
RP SEQUENCE, CHARACTERIZATION, AND SYNTHESIS.
RC STRAIN-FERUSSAC; TISSUE-Ganglion;
RX MDLINE-98273551; PubMed-2597281;
RA Kamatani Y., Minakata H., Kenny P.T.M., Iwashita T., Watanabe K.,
RA Punase K., Sun X.P., Yongsiri A., Kim K.H., Novales-Li P.,
RA Novales E.T., Kanapi C.G., Takeuchi H., Nomoto K.;
RT "Achatin-I, an endogenous neuroexcitatory tetrapeptide from Achatina
   fulica Ferrussac containing a D-amino acid residue."
RL Biochem. Biophys. Res. Commun. 160:1015-1020(1989).
RN [2]
RP CHARACTERIZATION.
RC STRAIN-FERUSSAC; TISSUE-Heart atrium;
RX MDLINE-91264856; PubMed-1675568;
RA Fujimoto K., Kubota I., Yasuda-Kamatani Y., Minakata H., Nomoto K.,
RA Iwashita M., Harada A., Muneoka Y., Kobayashi M.;
RT "Purification of achatin-I from the atria of the African giant snail,
   Achatina fulica, and its possible function."
RL Biochem. Biophys. Res. Commun. 177:847-853(1991).
RN [3]
RP X-RAY CRYSTALLOGRAPHY.
RX MDLINE-93014529; PubMed-1399265;
RA Ishida T., In Y., Doi M., Inoue M., Yasuda-Kamatani Y., Minakata H.,
RA Iwashita T., Nomoto K.;
RT "Crystal structure and molecular conformation of achatin-I
   (H-Gly-D-Phe-Ala-Asp-OH), an endogenous neuropeptide containing a
   D-amino acid residue."
RL Int. J. Pept. Protein Res. 39:258-264(1992).
CC -1- FUNCTION: NEUROEXCITATORY PEPTIDE; INCREASES THE IMPULSE FREQUENCY
CC AND PRODUCES A SPIKE BROADENING OF THE IDENTIFIED HEART EXCITATORY
CC NEURON (POM); ALSO ENHANCES THE AMPLITUDE AND FREQUENCY OF THE
CC HEART BEAT. HAS ALSO AN EFFECT ON SEVERAL OTHER MUSCLES.
DR PIR: A32480; A32480.
KM Hormone; D-amino acid.
FT MOD_RES 2
SQ SEQUENCE 4 AA; 408 MW; 6AADD9C810000000 CRC64;

Query Match 21.2%; Score 7; DB 1; Length 4;
Best Local Similarity 50.0%; Pred. No. 1e+05;
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 FS 5
DB 2 FA 3

RESULT 12
FMRF_MACNI STANDARD; PRT; 4 AA.
AC P01162;
DT 21-JUN-1986 (Rel. 01, Created)
DT 21-JUN-1986 (Rel. 01, Last sequence update)
DT 15-JUN-1998 (Rel. 36, Last annotation update)
DE FMRFamide (Peak C) (Cardioexcitatory neuropeptide).
OS Macrocallista nimbosa (Sun-ray clam),
OS Nerereis virens (Sandworm),
OS Hirudo medicinalis (medical leech), and
OS Helisoma trivolvis (Snail).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Heteroconchia; Veneroida;
OC Veneroidea; Veneridae; Macrocallista.
OX NCBI_TaxID=6594, 6353, 6421, 27815;
RN [1]
RP SEQUENCE, AND SYNTHESIS.
RC SPECIES-M. nimbosa; TISSUE-Cerebral pedal, and Visceral ganglion;
RX MDLINE-77215956; PubMed-877582;
RA Price D.A., Greenberg M.J.;

```

```

RT *structure of a molluscan cardioexcitatory neuropeptide.*;
RL Science 197;670-671(1977).
RN [2]
RN SPECIES-M.limbosa; TISSUE-ganglion;
RC MEDLINE-78012038; Pubmed-909875;
RA Price D.A., Greenberg M.J.;
RT "Purification and characterization of a cardioexcitatory neuropeptide
RL from the central ganglia of a bivalve mollusc.";
RN Prep. Biochem. 7;261-281(1977).
RN [3]
RN SPECIES-N.virens;
RC MEDLINE-90259866; Pubmed-2342992;
RA Krajinak K.G., Price D.A.;
RT "Authentic FMRFamide is present in the polychaete Nereis virens.";
RN Peptides 11;75-77(1990).
RN [4]
RN SPECIES-H.trivoltis; TISSUE-kidney;
RC MEDLINE-94286417; Pubmed-7912428;
RA Madrid K.P., Price D.A., Greenberg M.J., Khan H.R., Saleuddin A.S.M.;
RT "FMRFamide-related peptides from the kidney of the snail, Helisoma
RT trivoltis.";
RL Peptides 15;31-36(1994).
CC -I- FUNCTION: MYOACTIVE; CARDIOEXCITATORY SUBSTANCE. PHARMACOLOGICAL
CC ACTIVITIES INCLUDE AUGMENTATION, INDUCTION, AND REGULATION OF
CC CARDIAC CONTRACTION
CC -I- SIMILARITY: BELONGS TO THE FAMP (FMRFAMIDE RELATED PEPTIDE)
CC FAMILY.
DR PIR: A01426; ECKM.
DR PIR: A60418; A60418.
KW Neuropeptide; Amidation.
FT MOD.RES 4
SO SEQUENCE 4 AA; 600 MW; 69D40699A0000000 CRC64;

Query Match 21.2%; Score 7; DB 1; Length 4;
Best Local Similarity 33.3%; Pred. No. 1e+05; 1; Indels 0; Gaps 0;
Matches 1; Conservative 1; Mismatches 1;

QY 4 FSK 6
   1;
DB 1 FMR 3

RESULT 13
TUFT_HUMAN STANDARD; PRT; 4 AA.
ID TUFT_HUMAN
AC P01858;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phagocytosis-stimulating peptide (Tuftsin).
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SPECIES-H.
RC MEDLINE-72187087; Pubmed-4112769;
RA Nishio K., Constantinou A., Satoh P.S., Najjar V.A.;
RT "The characteristics, isolation and synthesis of the phagocytosis
RL stimulating peptide tuftsin.";
RN Biochem. Biophys. Res. Commun. 47;172-179(1972).
RN [2]

```

```

RP IMMUNOGLOBULIN CLASS.
RX MEDLINE-68091045; Pubmed-4169272;
RA Fidalgo B.V., Najjar V.A.;
RT "The physiological role of the lymphoid system, VI. The stimulatory
RT effect of leucophilic gamma globulin (leucokinin) on the phagocytic
RL activity of human polymorphonuclear leucocyte.";
RN Biochemistry 6;3386-3392(1967).
CC -I- MISCELLANEOUS: AN IGG (CALLED LEUCOKININ) BINDS REVERSIBLY TO THE
CC CELL MEMBRANE OF NEUTROPHILS IN THE BLOOD. LEUCOKININASE ON THE
CC MEMBRANE RELEASES THE ACTIVE PEPTIDE TUFTSIN FROM THE GAMMA CHAIN.
CC TUFTSIN IS ESSENTIAL FOR MAXIMUM STIMULATION OF THE PHAGOCYTIC
CC ACTIVITY OF NEUTROPHILS.
DR PIR: A02147; A02147.
DR MIM: 191150; -.
SO SEQUENCE 4 AA; 501 MW; 74176321C0000000 CRC64;

Query Match 21.2%; Score 7; DB 1; Length 4;
Best Local Similarity 50.0%; Pred. No. 1e+05;
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TR 2
   1;
DB 1 TK 2

RESULT 14
ID A14_CARMA
TUFT_HUMAN STANDARD; PRT; 5 AA.
AC P81817;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Carcinustatin 14.
OS Carcinus maenas (Common shore crab) (Green crab).
CC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
CC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
CC Eubrachyura; Portunidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RN SPECIES-H.
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;
RX MEDLINE-98121193; Pubmed-9461295;
RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA Thorpe A.;
RT "Isolation and identification of multiple neuropeptides of the
RT allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250;727-734(1997).
CC -I- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC -I- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
CC Neuropeptide; Amidation; Multigene family.
KW Neuropeptide; Amidation (POTENTIAL).
FT MOD.RES 5
SO SEQUENCE 5 AA; 586 MW; 672879D5A8300000 CRC64;

Query Match 21.2%; Score 7; DB 1; Length 5;
Best Local Similarity 50.0%; Pred. No. 1e+05;
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 FS 5
   1;
DB 1 YS 2

RESULT 15
RELL_LITRU STANDARD; PRT; 5 AA.
ID RELL_LITRU
AC P82070;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Rubellidin 1.1.
OS Litoria rubella (Desert tree frog).
OS

```

CC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
CC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litoria.  
OX NCBI\_TaxID-104895;  
RN [1]  
RP SEQUENCE, AND MASS SPECTROMETRY.  
RC TISSUE-Skin secretion;  
RA Stelbörner S.T., Weiboltz P.A., Waugh R.J., Bowle J.H., Gao C.,  
RT Tyler M.J., Wallace J.C.;  
RT "The structure of new peptides from the Australian red tree frog  
Litoria tubella: the skin peptide profile as a probe for the study  
of evolutionary trends of amphibians.";  
RL Aust. J. Chem. 49:955-963(1996).  
CC -1- FUNCTION: SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC  
CC ACTIVITY.  
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
CC -1- MASS SPECTROMETRY: MW=598; METHOD-FAB.  
KM Amphibian skin.  
SQ SEQUENCE 5 AA; 598 MW; 6DD9C9CAB2A00000 CRC64;

Query Match 21.28; Score 7; DB 1; Length 5;  
Best Local Similarity 50.08; Pred. No. 1e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 FS 5  
|:  
Db 4 FA 5

Search completed: July 15, 2002, 13:38:10  
Job time: 710 sec





DN 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DN 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE CHLOROPLAST 30S RIBOSOMAL PROTEIN S19 BETA (FRAGMENT).  
OS Spinacia oleracea (Spinach).  
OG Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllales; Caryophyllales; Chenopodiaceae; Spinacia.  
ON NCBI\_TaxID=3562;  
RN [1]  
RP SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.  
RC STRAIN=CV. ALVARO; TISSUE=LEAF;  
RX MEDLINE=20435797; PubMed=10874019;  
RA Yamauchi K., von Knebelbach K., Subramanian A.R.:  
RT "The plastid ribosomal proteins: identification of all the proteins in  
the small subunit of an organelle ribosome (chloroplast)."  
RL J. Biol. Chem. 37:28435-28465(2000).  
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA.  
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES.  
CC -1- MASS SPECTROMETRY: MW=10477.0; METHOD=ELECTROSPRAY.  
CC -1- MISCELLANEOUS: S19 ALPHA AND BETA FORMS DIFFER IN P.I. S19 BETA  
FORM IS THE MINOR BASIC FORM.  
CC -1- MISCELLANEOUS: ON THE 2D-GEL, ITS MW IS: 12 KDA.  
CC -1- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.  
DR InterPro: IPR002222; Ribosomal\_S19  
DR Pfam: PF00203; Ribosomal\_S19; PARTIAL.  
DR PRINTS: PR00975; RIBOSOMAL\_S19; PARTIAL.  
DR PROSITE: PS00323; RIBOSOMAL\_S19; PARTIAL.  
KW Ribosomal protein; Chloroplast; rRNA-binding.  
FT NON\_TER 6  
FT SEQUENCE 6 AA: 732 MW: 63333735A11C000 CRC64;

Query Match 39.4%; Score 13; DB 10; Length 6;  
Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 TRFSK 6  
DB 1 TRSLK 6

RESULT 3  
ID Q15903 PRELIMINARY; PRT; 7 AA.  
AC Q15903;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE (CLONE XP7E73) (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
ON NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RA Lee C.-C., Yezland A., Wehnert M., Bailey J., Couch L., Xiong M.,  
RA Coulbough M.T., Chinnault C.A., Baldwin A., Lindsay E.A., Zhao Z.-Y.,  
RA Caskey C.T.H.;  
RT "Isolation of chromosome-specific genes by reciprocal probing of  
RT arrayed cDNAs and cosmid libraries."  
RL Hum. Mol. Genet. 0:0-0(1995).  
DR EMBL: L32082; AAF73893.1; -.  
FT NON\_TER 1  
FT NON\_TER 7  
FT SEQUENCE 7 AA: 849 MW: 6B040339CDD3DB0 CRC64;

Query Match 27.3%; Score 9; DB 4; Length 7;  
Best Local Similarity 20.0%; Pred. No. 5.6e+05;

Matches 1; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 2 RFSK 6  
DB 2 KAFK 6

RESULT 4  
ID Q95945 PRELIMINARY; PRT; 7 AA.  
AC Q95945;  
DT 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE INSIDE INTRON 5 (FRAGMENT).  
OS Saccharomyces cerevisiae (Baker's yeast).  
OG Mitochondrion.  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
ON NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D273-10B;  
RX MEDLINE=1069885; PubMed=6254986;  
RA Bonitz S.G., Coruzzi G., Thalenfeld B., Tzagoloff A., Macino G.;  
RT "Assembly of the mitochondrial membrane system: Structure and  
RT nucleotide sequence of the gene coding for subunit 1 of yeast  
cytochrome oxidase."  
RL J. Biol. Chem. 255:11927-11941(1980).  
DR EMBL: V00694; CAA24066.1; -.  
KW Mitochondrion.  
FT NON\_TER 1  
FT SEQUENCE 7 AA: 859 MW: 75B7232362CD460 CRC64;

Query Match 27.3%; Score 9; DB 8; Length 7;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 KL 7  
DB 5 KL 6

RESULT 5  
ID Q9C5B3 PRELIMINARY; PRT; 7 AA.  
AC Q9C5B3;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYDROTHERGAL 0.7 KDA PROTEIN (FRAGMENT).  
GN D101.10A-2B.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
ON NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=ROOT;  
RX MEDLINE=21171025; PubMed=11277426;  
RA Vercuteren I., Van Der Schueren E., Van Montagu M., Gheysen G.;  
RT "Arabidopsis thaliana genes expressed in the early compatible  
RT interaction with root-knot nematodes."  
RL Mol. Plant Microbe Interact. 14:288-299(2001).  
DR EMBL: AJ286350; CAB71014.2; -.  
KW Hypothetical protein.  
FT NON\_TER 1  
FT NON\_TER 7  
FT SEQUENCE 7 AA: 719 MW: 6732C7287EB325D0 CRC64;

Query Match 27.3%; Score 9; DB 10; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 5 SK 6  
 1 1  
 Db 1 SK 2

RESULT 6  
 P82445 PRELIMINARY; PRT: 7 AA.  
 ID P82445:  
 AC P82445:  
 DT 01-JUN-2000 (TREMblrel. 14, Created)  
 DT 01-JUN-2000 (TREMblrel. 14, Last sequence update)  
 DE 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
 DE 10 KDA CELL WALL PROTEIN (FRAGMENT).  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=CV. PETIT HAVANA;  
 RA Blee K.A., Bonham V.A., Mitchell G.P., Robertson D., Stabbs A.R.,  
 RA Mojtaszek P., Bolwell G.P.,  
 RT Proteomic study of secondary cell wall proteins from transformed  
 RT tobacco culture.\*;  
 RL Planta 0:0-0(2000).  
 CC -1- SUBCELLULAR LOCATION: CELL WALL.  
 CC -1- TISSUE SPECIFICITY: XYLEM.  
 KW Cell wall.  
 KM NON\_TER  
 FT SEQUENCE 7 AA; 758 MW; 69D2C1B62D1B2A0 CRC64;  
 SQ

Query Match 27.3%; Score 9; DB 10; Length 7;  
 Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
 Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 3 IF 4  
 1 1  
 Db 6 VF 7

RESULT 7  
 O66205 PRELIMINARY; PRT: 7 AA.  
 ID O66205:  
 AC O66205:  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE MEMBRANE PROTEIN (1 IS 3RD BASE IN CODON) (FRAGMENT).  
 OS Transmissible gastroenteritis virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=11149;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=F5772/70;  
 RX MEDLINE=88216185; PubMed=2835592;  
 RA Britton P., Carnes R.S., Page K.W., Garves D.J., Parra F.,  
 RT Sequence of the Nucleoprotein Gene from a Virulent British Field  
 RT Isolate of Transmissible Gastroenteritis Virus and Its Expression in  
 RT Saccharomyces Cerevisiae.\*;  
 RL Mol. Microbiol. 2:89-99(1988).  
 DR EMBL: Y00542; CAA68606.1; -;  
 FT NON\_TER  
 SQ SEQUENCE 7 AA; 869 MW; 72C691E727233B70 CRC64;  
 FT

Query Match 27.3%; Score 9; DB 12; Length 7;

Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 6 KL 7  
 1 1  
 Db 2 KL 3

RESULT 8  
 O99182 PRELIMINARY; PRT: 7 AA.  
 ID O99182:  
 AC O99182:  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE CYTOCHROME OXIDASE I (FRAGMENT).  
 GN COI.  
 OS Gnatholebias zonatus.  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;  
 OC Cyprinodontiformes; Aplocheilidae; Rivulinae; Gnatholebias.  
 OX NCBI\_TaxID=135316;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=20072928; PubMed=10603257;  
 RA Murphy W.J., Jomerson J.E., Collier G.E.,  
 RA "Phylogeny of the Neotropical Killifish Family Rivulidae  
 RT (Cyprinodontiformes, Aplocheilidae) Inferred from mitochondrial DNA  
 RT sequences.\*";  
 RL Mol. Phylogenet. Evol. 13:289-301(1999).  
 DR EMBL: AF002591; AAD01074.1; -;  
 KM Mitochondrion.  
 FT NON\_TER  
 SQ SEQUENCE 7 AA; 899 MW; 672721F6CB572030 CRC64;  
 FT

Query Match 24.2%; Score 8; DB 8; Length 7;  
 Best Local Similarity 20.0%; Pred. No. 5.6e+05;  
 Matches 1; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 OY 3 IESKL 7  
 1 1  
 Db 2 LYQHL 6

RESULT 9  
 O98866 PRELIMINARY; PRT: 7 AA.  
 ID O98866:  
 AC O98866:  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DE 01-MAY-1999 (TREMblrel. 10, Last annotation update)  
 DE CYTOCHROME B/P SUBUNIT IV (FRAGMENT).  
 OS Spinacia oleracea (Spinach).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
 OX NCBI\_TaxID=3562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=86120353; PubMed=3003688.  
 RA Sijben-Mueller G., Halliack R.B., Alt J., Westhoff P., Herrmann R.G.,  
 RT "Spinach plastid genes coding for initiation factor IF-1, ribosomal  
 RT protein S11 and RNA polymerase alpha-subunit.\*";  
 RL Nucleic Acids Res. 14:1029-1044(1986).  
 DR EMBL: X03496; CAA7215.1; -;  
 KM Chloroplast.  
 FT NON\_TER  
 SQ SEQUENCE 7 AA; 907 MW; 644729D77409CA20 CRC64;  
 FT

```

Query Match      24.2%; Score 8; DB 8; Length 7;
Best Local Similarity 66.7%; Pred. No. 5.6e+05;
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 RIF 4
DB 3 RPF 5

RESULT 10
P82070 PRELIMINARY; PRT; 5 AA.
AC P82070;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE RUBELLIDIN 1.1.
OS Litorea rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Litorea.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE-SKIN SECRETION;
RA Steinboerner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,
RA Tyler M.J., Wallace J.C.;
RT 'The structure of new peptides from the Australian red tree frog
RT 'Litorea rubella', the skin peptide profile as a probe for the study
RT of evolutionary trends of amphibians.';
RL Aust. J. Chem. 49:955-963(1996).
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR
CC ANTIBIOTIC ACTIVITY.
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.
CC -1- MASS SPECTROMETRY: MW=598; METHOD=FAB.
KW Amphibian skin.
SQ SEQUENCE 5 AA: 598 MW: 6DD9C9CAB2A00000 CRC64;

Query Match      21.2%; Score 7; DB 13; Length 5;
Best Local Similarity 50.0%; Pred. No. 5.6e+05;
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 FS 5
DB 4 FA 5

RESULT 11
P82071 PRELIMINARY; PRT; 5 AA.
AC P82071;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE RUBELLIDIN 2.1.
OS Litorea rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Litorea.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE-SKIN SECRETION;
RA Steinboerner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,
RA Tyler M.J., Wallace J.C.;
RT 'The structure of new peptides from the Australian red tree frog
RT 'Litorea rubella', the skin peptide profile as a probe for the study
RT of evolutionary trends of amphibians.';
RL Aust. J. Chem. 49:955-963(1996).
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR
CC ANTIBIOTIC ACTIVITY.
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.
CC -1- MASS SPECTROMETRY: MW=598; METHOD=FAB.
KW Amphibian skin.
SQ SEQUENCE 5 AA: 598 MW: 6DD9C9CAB2A00000 CRC64;

Query Match      21.2%; Score 7; DB 13; Length 5;
Best Local Similarity 50.0%; Pred. No. 5.6e+05;
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 FS 5
DB 4 FA 5

RESULT 12
P82072 PRELIMINARY; PRT; 5 AA.
AC P82072;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE RUBELLIDIN 3.1.
OS Litorea rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Litorea.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE-SKIN SECRETION;
RA Steinboerner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,
RA Tyler M.J., Wallace J.C.;
RT 'The structure of new peptides from the Australian red tree frog
RT 'Litorea rubella', the skin peptide profile as a probe for the study
RT of evolutionary trends of amphibians.';
RL Aust. J. Chem. 49:955-963(1996).
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR
CC ANTIBIOTIC ACTIVITY.
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.
CC -1- MASS SPECTROMETRY: MW=655; METHOD=FAB.
KW Amphibian skin; Amidation.
PT MOD_RES 5
SQ SEQUENCE 5 AA: 656 MW: 71A9C9CB10300000 CRC64;

Query Match      21.2%; Score 7; DB 13; Length 5;
Best Local Similarity 50.0%; Pred. No. 5.6e+05;
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 FS 5
DB 4 FT 5

RESULT 13
P82073 PRELIMINARY; PRT; 5 AA.
AC P82073;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE RUBELLIDIN 3.2.
OS Litorea rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Litorea.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE.
RC TISSUE-SKIN SECRETION;
RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;
RT 'Peptides from the skin glands of the Australian buzzing tree frog

```

RT Litoria electrica. Comparison with the skin peptides from Litoria  
rubella.\*;  
RL Aust. J. Chem. 52:0-0(1999).  
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR  
CC ANTI-BIOTIC ACTIVITY.  
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
KW Amphibian skin.  
SQ SEQUENCE 5 AA: 570 MW: 719C9C862A00000 CRC64;

Query Match 21.2%; Score 7; DB 13; Length 5;  
Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 FS 5  
1:  
DB 4 FT 5

RESULT 14  
P82181  
ID P82181; PRELIMINARY; PRT; 6 AA.  
AC P82181; 01-JUN-2000 (TREMBLrel. 14, Created)  
DT 01-JUN-2000 (TREMBLrel. 14, Last sequence update)  
DE CHLOROPLAST 50S RIBOSOMAL PROTEIN L10 BETA (FRAGMENT).  
OS Spinacia oleracea (Spinach).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
OX NCBI\_TaxID=3562;  
RN [1]  
RP SEQUENCE.  
RC STRAIN-CV. ALVARO; TISSUE-LEAF;  
RA MEDLINE=20435798; PubMed=10874046;  
RX Yamaguchi K., Subramanian A.R.;  
RT "The plastid ribosomal proteins. Identification of all the proteins in  
RT the 50 S subunit of an organelle ribosome (chloroplast).";  
RL J. Biol. Chem. 275:28466-28482(2000).  
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.  
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES.  
CC -1- MISCELLANEOUS: ON THE 2D-GEL ITS MW IS: 16.5 KDA.  
CC -1- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.  
DR InterPro: IPR001790; Ribosomal\_L10.  
DR pfam: PF00466; Ribosomal\_L10\_eub.  
DR PROSITE: PS01109; RIBOSOMAL\_L10; PARTIAL.  
KW Ribosomal protein; Chloroplast; rRNA-binding.  
FT NON\_TER 6  
SQ SEQUENCE 6 AA: 675 MW: 6321B415B05DB000 CRC64;

Query Match 21.2%; Score 7; DB 10; Length 6;  
Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TR 2  
1:  
DB 5 TK 6

RESULT 15  
P82182  
ID P82182; PRELIMINARY; PRT; 6 AA.  
AC P82182; 01-JUN-2000 (TREMBLrel. 14, Created)  
DT 01-JUN-2000 (TREMBLrel. 14, Last sequence update)  
DE CHLOROPLAST 50S RIBOSOMAL PROTEIN L10 GAMMA (FRAGMENT).  
OS Spinacia oleracea (Spinach).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
OX NCBI\_TaxID=3562;  
RN [1]  
RP SEQUENCE.  
RC STRAIN-CV. ALVARO; TISSUE-LEAF;  
RA MEDLINE=20435798; PubMed=10874046;  
RX Yamaguchi K., Subramanian A.R.;  
RT "The plastid ribosomal proteins. Identification of all the proteins in  
RT the 50 S subunit of an organelle ribosome (chloroplast).";  
RL J. Biol. Chem. 275:28466-28482(2000).  
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.  
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES.  
CC -1- MISCELLANEOUS: ON THE 2D-GEL ITS MW IS: 16.5 KDA.  
DR InterPro: IPR001790; Ribosomal\_L10.  
DR pfam: PF00466; Ribosomal\_L10\_eub.  
DR PROSITE: PS01109; RIBOSOMAL\_L10; PARTIAL.  
KW Ribosomal protein; Chloroplast; rRNA-binding.  
FT NON\_TER 6  
SQ SEQUENCE 6 AA: 675 MW: 6321B415B05DB000 CRC64;

Query Match 21.2%; Score 7; DB 10; Length 6;  
Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TR 2  
1:  
DB 5 TK 6

Search completed: July 15, 2002, 13:37:39  
Job time: 729 sec



Result No.	Score	Query Match	length	DB	ID	Description
1	23	69.7	7	20	AAW92345	Latent fluorescent
2	20	60.6	7	16	AAH87299	Plasmidium falcipar
3	20	60.6	7	21	AAH28929	Geminitivirus Rep pe
4	19	57.6	7	18	AAH18611	Aged band 3 peptid
5	18	54.5	4	21	AAH13882	L2/HN1 epitope m
6	18	54.5	6	17	AAH08723	Trypsin inhibitory
7	18	54.5	6	17	AAH08724	Trypsin inhibitory
8	18	54.5	6	19	AAH68419	Peptide #1 from cy
9	18	54.5	6	21	AAH15880	Consensus sequence
10	18	54.5	7	21	AAH13881	L2/HN1 epitope m
11	18	54.5	7	22	AAH98766	Human cell death F

12	AA018503	5	18	AA018503	Amino-terminal pep
13	AA068421	6	19	AA068421	Peptide #3 from cy
14	AA068428	6	19	AA068428	Peptide #10 from c
15	AA017746	6	20	AA017746	Calreticulin bindi
16	AA024501	7	13	AA024501	Monomeric repeat
17	AA040797	7	20	AA040797	Monomeric repeat
18	AA030137	7	20	AA030137	Amino acid loop 13
19	AA022771	7	21	AA022771	Scarfoid protein S
20	AA040463	4	5	AA040463	Scarfoid protein S
21	AA085776	4	17	AA085776	IMU receptor targ
22	AA050125	4	17	AA050125	N-terminal polynyl
23	AA087895	5	16	AA087895	Tetrapeptide usetu
24	AA012483	5	18	AA012483	Bovine, ovine and
25	AA006421	5	21	AA006421	Peptide inhibitor
26	AA006421	5	21	AA006421	Interleukin-6 anta
27	AA050143	6	15	AA050143	Claudin-1 cell adh
28	AA089186	6	16	AA089186	Ovine growth hormo
29	AA075336	6	19	AA075336	Ovine growth hormo
30	AA075268	6	19	AA075268	Ovine growth hormo
31	AA056849	6	19	AA056849	3' end of alkaline
32	AA064422	6	21	AA064422	Hexapeptide #1 bin
33	AA006424	6	21	AA006424	Hexapeptide #1 bin
34	AA006540	6	21	AA006540	Enzyme inhibitor P
35	AA006540	6	21	AA006540	Claudin-1 cell adh
36	AA006540	6	21	AA006540	Claudin-1 cell adh
37	AA006540	6	21	AA006540	Claudin-1 cell adh
38	AA006540	6	21	AA006540	Claudin-1 cell adh
39	AA006540	6	21	AA006540	Claudin-1 cell adh
40	AA006540	6	21	AA006540	Claudin-1 cell adh
41	AA006540	6	21	AA006540	Claudin-1 cell adh
42	AA006540	6	21	AA006540	Claudin-1 cell adh
43	AA006540	6	21	AA006540	Claudin-1 cell adh
44	AA006540	6	21	AA006540	Claudin-1 cell adh
45	AA006540	6	21	AA006540	Claudin-1 cell adh
46	AA006540	6	21	AA006540	Claudin-1 cell adh
47	AA006540	6	21	AA006540	Claudin-1 cell adh
48	AA006540	6	21	AA006540	Claudin-1 cell adh
49	AA006540	6	21	AA006540	Claudin-1 cell adh
50	AA006540	6	21	AA006540	Claudin-1 cell adh
51	AA006540	6	21	AA006540	Claudin-1 cell adh
52	AA006540	6	21	AA006540	Claudin-1 cell adh
53	AA006540	6	21	AA006540	Claudin-1 cell adh
54	AA006540	6	21	AA006540	Claudin-1 cell adh
55	AA006540	6	21	AA006540	Claudin-1 cell adh
56	AA006540	6	21	AA006540	Claudin-1 cell adh
57	AA006540	6	21	AA006540	Claudin-1 cell adh
58	AA006540	6	21	AA006540	Claudin-1 cell adh
59	AA006540	6	21	AA006540	Claudin-1 cell adh
60	AA006540	6	21	AA006540	Claudin-1 cell adh
61	AA006540	6	21	AA006540	Claudin-1 cell adh
62	AA006540	6	21	AA006540	Claudin-1 cell adh
63	AA006540	6	21	AA006540	Claudin-1 cell adh
64	AA006540	6	21	AA006540	Claudin-1 cell adh
65	AA006540	6	21	AA006540	Claudin-1 cell adh
66	AA006540	6	21	AA006540	Claudin-1 cell adh
67	AA006540	6	21	AA006540	Claudin-1 cell adh
68	AA006540	6	21	AA006540	Claudin-1 cell adh
69	AA006540	6	21	AA006540	Claudin-1 cell adh
70	AA006540	6	21	AA006540	Claudin-1 cell adh
71	AA006540	6	21	AA006540	Claudin-1 cell adh
72	AA006540	6	21	AA006540	Claudin-1 cell adh
73	AA006540	6	21	AA006540	Claudin-1 cell adh
74	AA006540	6	21	AA006540	Claudin-1 cell adh
75	AA006540	6	21	AA006540	Claudin-1 cell adh
76	AA006540	6	21	AA006540	Claudin-1 cell adh
77	AA006540	6	21	AA006540	Claudin-1 cell adh
78	AA006540	6	21	AA006540	Claudin-1 cell adh
79	AA006540	6	21	AA006540	Claudin-1 cell adh
80	AA006540	6	21	AA006540	Claudin-1 cell adh
81	AA006540	6	21	AA006540	Claudin-1 cell adh
82	AA006540	6	21	AA006540	Claudin-1 cell adh
83	AA006540	6	21	AA006540	Claudin-1 cell adh
84	AA006540	6	21	AA006540	Claudin-1 cell adh
85	AA006540	6	21	AA006540	Claudin-1 cell adh
86	AA006540	6	21	AA006540	Claudin-1 cell adh
87	AA006540	6	21	AA006540	Claudin-1 cell adh
88	AA006540	6	21	AA006540	Claudin-1 cell adh
89	AA006540	6	21	AA006540	Claudin-1 cell adh
90	AA006540	6	21	AA006540	Claudin-1 cell adh
91	AA006540	6	21	AA006540	Claudin-1 cell adh
92	AA006540	6	21	AA006540	Claudin-1 cell adh
93	AA006540	6	21	AA006540	Claudin-1 cell adh
94	AA006540	6	21	AA006540	Claudin-1 cell adh
95	AA006540	6	21	AA006540	Claudin-1 cell adh
96	AA006540	6	21	AA006540	Claudin-1 cell adh
97	AA006540	6	21	AA006540	Claudin-1 cell adh
98	AA006540	6	21	AA006540	Claudin-1 cell adh
99	AA006540	6	21	AA006540	Claudin-1 cell adh
100	AA006540	6	21	AA006540	Claudin-1 cell adh

RESULT	1
AAW92345	
ID	AAW92345 standard; peptide; 7 AA.
XX	
AC	AAW92345;
XX	
DT	09-APR-1999 (first entry).
XX	
DE	Latent fluorescent peptide #9.
XX	
KW	Fluorescent peptide; masking group; fluorescent group; sensitive; stable;
XX	
KW	phosphatase activity; phosphoric acid.
XX	
OS	Synthetic.
XX	
PN	JP11012297-A.
XX	
PD	19-JAN-1999.
XX	
PF	21-APR-1998; 98JP-0126684.
XX	
PR	30-APR-1997; 97JP-0126463.
XX	
PA	(NISHU) NISHIKATA S.
XX	
DR	WPI; 1999-148570/13.
XX	
PT	New latent fluorescent peptide(s) - useful for determining protein
XX	
PT	phosphatase activity
XX	
PS	Claim 5; Page 9; 11pp; Japanese.
XX	
CC	AAW92337-W92350 are novel latent fluorescent (optionally protected)
XX	
CC	peptides having a self fluorescent group and masking group in the

CC molecule and phosphoric acid groups in the masking group and between the  
 CC fluorescent and masking groups. The invention also describes peptide  
 CC derivatives of formula (A1-X-A2), especially composed of 3-42 amino  
 CC acids; X = amino acid residue comprising phosphoric acid residue; A1,  
 CC A2 = (optionally protected) amino acid residue, (optionally protected)  
 CC peptide chain, one comprises a masking group at the terminal or side  
 CC chain and the other has a self fluorescent group capable of being masked  
 CC with the intramolecular masking group. The peptides are useful for  
 CC determining protein phosphatase activity by the changes of intensity of  
 CC fluorescence. The fluorescent peptides are stable reagents for sensitive  
 CC and rapid determination of protein phosphatase activity in crude sample  
 CC even contaminated sample containing phosphoric acid ion.  
 XX  
 SQ Sequence 7 AA;

Query Match 60.7%; Score 23; DB 20; Length 7;  
 Best Local Similarity 80.0%; Pred. No. 6.4e+05;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 RIFSK 6  
 Db 2 rfsk 6

RESULT 2

ID AAR87299 standard; peptide: 7 AA.

XX AAR87299;

DT 16-MAY-1996 (first entry)

DE Plasmodium falciparum HLA-B17 epitope 1555.

XX Liver stage antigen; LSA-1; human leucocyte antigen; HLA; class I;

KW HLA-B17; epitope; malaria; vaccine; CTL induction;

KM cytotoxic T lymphocyte.

XX Plasmodium falciparum.

PN W09526982-A2.

PD 12-OCT-1995.

XX 31-MAR-1995; 95MO-GB00727.

PF 31-MAR-1994; 94GB-0006492.

XX (ISIS-) ISIS INNOVATION LTD.

PI Aideo M, Allsopp CEM, Hill AVS, Lalvani A, Piebanski M;

PI Whittle HC;

DR WPI; 1995-358584/46.

XX Claim 1; Page 19; 23pp; English.

CC Cytotoxic T lymphocytes from malaria-exposed Gambian individuals  
 CC with HLA-B17 showed significant lysis of a large pool of peptides  
 CC (AAR87287-R87299) derived from Plasmodium falciparum liver stage  
 CC antigen-1. These peptides will be useful in a malaria vaccine.  
 XX  
 SQ Sequence 7 AA;

Query Match 60.6%; Score 20; DB 16; Length 7;  
 Best Local Similarity 57.1%; Pred. No. 6.4e+05;  
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 RIFSKL 7  
 Db 1 tkylmkl 7

RESULT 3

ID AAB28929 standard; peptide: 7 AA.

XX AAB28929;

DT 25-JAN-2001 (first entry)

DE Geminivirus Rep peptide sequence #31.

XX Geminivirus; replication associated protein; rep; iteron;

KW antagonists; plant.

OS Geminivirus.

PN W0200043494-A2.

PD 27-JUL-2000.

PF 27-JAN-2000; 2000MO-US01849.

PR 26-JAN-1999; 99US-0117285.

XX (SCRI) SCRIPPS RES INST.

PI Fauquet C, Chatterji A;

DR WPI; 2000-499224/44.

XX Producing plants resistant to geminivirus, and inhibiting geminivirus

PT replication in plants, by introducing replication associated protein

PT iteron antagonists into the plant, plant cells or propagules -

XX Claim 5; Page 106; 172pp; English.

CC The present invention relates to methods for producing plants resistant  
 CC to geminivirus, involving introducing a geminivirus replication  
 CC associated protein (Rep)-iteron antagonist into a plant. The antagonist  
 CC is a nucleotide sequence of a geminivirus iteron capable of binding to  
 CC a Rep protein or a defective Rep which has a conserved geminivirus  
 CC iteron binding site. The present sequence is a geminivirus Rep peptide  
 CC sequence.

SQ Sequence 7 AA;

Query Match 60.6%; Score 20; DB 21; Length 7;  
 Best Local Similarity 60.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 RIFSK 6  
 Db 2 rvyek 6

RESULT 4

ID AAW18611 standard; peptide: 7 AA.

XX AAW18611;

DT 04-MAR-1998 (first entry)

XX Aged band 3 peptide (residues 534-540) epitope.

DE Band 3 protein; antibody; aging antigenic site; Alzheimer's disease;  
 KW phosphorylation; detection; epitope.  
 XX



OS Homo sapiens.  
 XX  
 PN WO9726537-A1.  
 XX  
 PD 24-JUL-1997.  
 XX  
 PF 13-DEC-1996; 96MO-US20465.  
 XX  
 PR 19-JAN-1996; 96US-0010250.  
 XX  
 PA (RESE ) RESEARCH CORP TECHNOLOGIES INC.  
 XX  
 PI Kay MMB;  
 XX  
 DR WPI: 1997-385478/35.  
 XX  
 PT Detecting Alzheimer's disease using antibody that recognises aged  
 PT band 3 protein in tissues - or from reduced degree of band 3 protein  
 XX phosphorylation, can be applied to blood or brain samples  
 XX  
 PS Disclosure; Page 12; 45pp; English.  
 XX  
 CC This is an aging antigenic band 3 peptide (residues 534-540) to which a  
 CC specific antibody can bind to. Band 3 is a ubiquitous anion-exchange  
 CC protein and ages as cells and tissues age. Antibodies have been developed  
 CC against this aged band 3. These antibodies bind to distinct regions of  
 CC band 3 in old cells (aging antigenic sites) but not middle aged or young  
 CC cells. This can be used for detecting Alzheimer's disease. A tissue  
 CC sample containing band 3 from a patient suspected of having Alzheimer's  
 CC disease is treated with an antibody that can differentiate between the  
 CC Alzheimer's (aged) and normal band 3, under complex-forming conditions  
 CC and detecting any complex formed. A tissue containing band 3 from a  
 CC healthy control is treated in a similar manner and the amounts of complex  
 CC formed are compared. A significantly greater formation of complex in the  
 CC suspect sample as compared with that of the control is indicative of  
 CC Alzheimer's disease. The disease can also be detected by comparing the  
 CC degrees of phosphorylation of band 3 or its degradation products in  
 CC suspect and control samples. A significant decrease in phosphorylation  
 CC in the suspect sample indicates Alzheimer's disease.  
 CC  
 SQ Sequence 7 AA:  
 Query Match 57.6%; Score 19; DB 18; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 PSKL 7  
 Db 4 fskl 7  
 DE  
 DE L2/HNK1 epitope mimic peptide homology sequence #2.  
 XX  
 KM L2/HNK1: epitope; nervous tissue; apoptosis; necrosis;  
 KM Parkinson's disease; multiple sclerosis; spinal cord injury;  
 KM Alzheimer's disease; dementia.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200050447-A1.  
 XX  
 PD 31-AUG-2000.  
 XX  
 PR 24-FEB-2000; 2000MO-US04730.

XX  
 PR 24-FEB-1999; 99US-0121327.  
 PR 24-FEB-1999; 99US-0256970.  
 PR 23-SEP-1999; 99US-0155492.  
 PR 23-SEP-1999; 99US-0404431.  
 PR 23-FEB-2000; 2000US-0511956.  
 XX  
 PA (ACOR-) ACORDA THERAPEUTICS.  
 XX  
 PI Schachner M, Neuberger TV, Herzberg U, Simon M;  
 XX  
 DR WPI: 2000-579163/54.  
 XX  
 PT New isolated peptide which mimics a carbohydrate epitope is useful for  
 PT neuroprotection  
 XX  
 PS Claim 9; Page 124; 214pp; English.  
 XX  
 CC The present invention relates to L2/HNK1 carbohydrate epitope mimic  
 CC peptides. The L2/HNK1 epitope is predominantly expressed on glycolipids  
 CC and glycoproteins from nervous tissue. Peptides that mimic the L2/HNK1  
 CC epitope were isolated by screening phage peptide display libraries with  
 CC antibodies to L2/HNK1. The present sequence is the homologous region  
 CC between consensus sequences of mimic peptides isolated by different  
 CC antibodies. The peptides may be used for enhancing memory and for  
 CC treating Alzheimer's disease or dementia. They may also be used for  
 CC and acute and chronic spinal cord injury.  
 CC  
 SQ Sequence 4 AA:  
 Query Match 54.5%; Score 18; DB 21; Length 4;  
 Best Local Similarity 75.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TRIF 4  
 Db 1 trlf 4  
 DE  
 DE Trypsin inhibitory peptide #4.  
 XX  
 DE  
 DE Lysand; synthetic combinatorial peptide library; hexamer; antibody;  
 KM antigen; receptor; inhibitor; trypsin.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FH Modified-site 1  
 FT /label= OTHER  
 FT /note= "Ac-Met"  
 FT Modified-site 6  
 FT /label= OTHER  
 FT /note= "Thr-NH2"  
 XX  
 PN US556762-A.  
 XX  
 PD 17-SEP-1996.  
 XX  
 PF 21-NOV-1990; 90US-0617023.  
 XX  
 PR 11-SEP-1992; 92US-0943709.  
 PR 21-NOV-1990; 90US-0617023.  
 PR 16-MAY-1991; 91US-0701656.

```
PR 19-NOV-1991. 91US-0797551.
XX
XX (HOUG-) HOUGHTEN PHARM INC.
XX
XX Appel JR, Houghten RA, Pinilla C;
XX
XX WPI; 1996-432985/43.
XX
XX Identifying oligopeptide ligands for an acceptor - by scanning
XX synthetic peptide combinatorial libraries comprising
XX self-solubilising, unsupported mixed oligopeptide(s)
XX
XX
XX Example 8; Column 59; 75pp; English.
XX
XX The invention relates to a method of identifying oligopeptide ligands to
XX a protein by scanning synthetic combinatorial peptide libraries (SCPL).
XX The SCPL comprise sets of hexamers which contain either one of 6
XX predetermined amino acids at one predetermined position in the hexamer
XX and each set may have one predetermined amino acid at 1 of 6
XX predetermined pos. in the hexamer. The method is useful for identifying
XX biologically active sequences of e.g. pharmaceutical use. The peptides
XX esp. inhibit the binding of e.g. an antibody to its antigen, a ligand to
XX its receptor, etc.
XX The peptides AAW08720-25 are N-terminally acylated and C-terminally
XX amidated peptides, isolated from a SCPL, which inhibit trypsin.
XX This peptide inhibited trypsin with an IC50 of 133 micromole.
XX
XX Sequence 6 AA:
SQ
XX
XX Query Match 54.5%; Score 18; DB 17; Length 6;
XX Best Local Similarity 60.0%; Pred. No. 6.4e+05;
XX Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 1 TRIPS 5
OY 1:11:
OY 2 tkft 6
DB
XX
XX RESULT 7
XX AAW08724
XX ID AAW08724 standard; peptide: 6 AA.
XX
XX AC AAW08724:
XX
XX DT 28-FEB-1997 (first entry)
XX
XX DE Trypsin inhibitory peptide #5.
XX
XX KW ligand; synthetic combinatorial peptide library; hexamer; antibody;
XX antigen; receptor; inhibitor; trypsin.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX Modified-site 1
XX FT /label= OTHER
XX FT /note= "Ac-Thr"
XX FT Modified-site 6
XX FT /label= OTHER
XX FT /note= "Thr-NH2"
XX
XX PN US5556762-A.
XX
XX PD 17-SEP-1996.
XX
XX PF 21-NOV-1990; 90US-0617023.
XX
XX PR 11-SEP-1992; 92US-0943709.
XX PR 21-NOV-1990; 90US-0617023.
XX PR 16-MAY-1991; 91US-0701658.
XX PR 19-NOV-1991; 91US-0797551.
XX
```

```
PA (HOUG-) HOUGHTEN PHARM INC.
XX
XX Appel JR, Houghten RA, Pinilla C;
XX
XX WPI; 1996-432985/43.
XX
XX Identifying oligopeptide ligands for an acceptor - by scanning
XX synthetic peptide combinatorial libraries comprising
XX self-solubilising, unsupported mixed oligopeptide(s)
XX
XX
XX Example 8; Column 59; 75pp; English.
XX
XX The invention relates to a method of identifying oligopeptide ligands to
XX a protein by scanning synthetic combinatorial peptide libraries (SCPL).
XX The SCPL comprise sets of hexamers which contain either one of 6
XX predetermined amino acids at one predetermined position in the hexamer
XX and each set may have one predetermined amino acid at 1 of 6
XX predetermined pos. in the hexamer. The method is useful for identifying
XX biologically active sequences of e.g. pharmaceutical use. The peptides
XX esp. inhibit the binding of e.g. an antibody to its antigen, a ligand to
XX its receptor, etc.
XX The peptides AAW08720-25 are N-terminally acylated and C-terminally
XX amidated peptides, isolated from a SCPL, which inhibit trypsin.
XX This peptide inhibited trypsin with an IC50 of 164 micromole.
XX
XX Sequence 6 AA:
SQ
XX
XX Query Match 54.5%; Score 18; DB 17; Length 6;
XX Best Local Similarity 60.0%; Pred. No. 6.4e+05;
XX Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 1 TRIPS 5
OY 1:11:
OY 2 tkft 6
DB
XX
XX RESULT 8
XX AAW68419
XX ID AAW68419 standard; peptide: 6 AA.
XX
XX AC AAW68419;
XX
XX DT 23-NOV-1998 (first entry)
XX
XX DE Peptide #1 from cyclic screening method for physiological activity.
XX
XX KW Screening method; physiological activity; hexapeptide; algorithm;
XX crossover; mutation; enzyme inhibitor; drug; foodstuff.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX Modified-site 1
XX FT /note= "acylated N-terminus"
XX FT Modified-site 6
XX FT /note= "amidated C-terminus"
XX
XX PN W09804580-A1.
XX
XX PD 05-FEB-1998.
XX
XX PF 23-JUL-1997; 97WO-0P02535.
XX
XX PR 26-JUL-1996; 96JP-0198096.
XX
XX PA (YAMA ) YAMANOUCHI PHARM CO LTD.
XX
XX PI Karube M, Yokobayashi Y;
XX
XX DR WPI; 1998-130616/12.
XX
XX Method for screening peptide(s) with high physiological activity -
PT
```

PT by determining activity, dropping least active, changing sequence in  
PT remaining peptide(s) and cycling through this sequence to obtain  
XX peptide(s) with high activity  
PS Disclosure; Page 9; 19pp; Japanese.  
XX  
CC Peptides AAW68419-W68434 represent peptides obtained by the cyclic  
CC screening method of the invention. The method comprises determining  
CC the level of a desired physiological activity in a series of  
CC hexapeptides and selecting those with the highest activities. The  
CC peptides are then processed using a genetic algorithm computer program  
CC which: (i) exchanges one or more amino acid residues between pairs of  
CC peptides (crossover) and/or (ii) substitutes amino acid residues in a  
CC peptide for different ones (mutation) (preferably with a mutation  
CC frequency of about 3% of the total number of residues). The altered  
CC peptides are synthesized and the screening cycle is repeated until  
CC peptides with the desired level of activity are generated. Peptides with  
CC a high activity (e.g. as enzyme inhibitors) can be used to produce  
CC compositions for use in the drug, foodstuff and other industries.  
XX  
SQ Sequence 6 AA:  
  
Query Match 54.5%; Score 18; DB 19; Length 6;  
Best Local Similarity 60.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 TRIFS 5  
|:|:  
Db 2 k1if 6  
  
RESULT 9  
AAB13880  
ID AAB13880 standard; peptide; 6 AA.  
XX  
AC AAB13880;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE Consensus sequence for mimic peptides selected by HNK1 antibody.  
XX  
KW L2/HNK1: epitope; nervous tissue; apoptosis; necrosis; antibody;  
KW Parkinson's disease; multiple sclerosis; spinal cord injury;  
XX Alzheimer's disease; dementia.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 6 /label= Val, Phe  
XX  
PN WO200050447-A1.  
XX  
PD 31-AUG-2000.  
XX  
PF 24-FEB-2000; 2000MO-US04730.  
XX  
PR 24-FEB-1999; 99US-0121327.  
PR 24-FEB-1999; 99US-0256970.  
PR 23-SEP-1999; 99US-0155452.  
PR 23-SEP-1999; 99US-040431.  
PR 23-FEB-2000; 2000US-0511956.  
XX  
PA (ACOR-) ACORDA THERAPEUTICS.  
XX  
PI Schachner M, Neuberger TJ, Herzberg U, Simon M;  
XX WPI; 2000-579163/54.  
XX  
PT New isolated peptide which mimics a carbohydrate epitope is useful for  
PT neuroprotection  
XX

PS Claim 7; Page 124; 214pp; English.  
XX  
CC The present invention relates to L2/HNK1 carbohydrate epitope mimic  
CC peptides. The L2/HNK1 epitope is predominantly expressed on glycolipids  
CC and glycoproteins from nervous tissue. Peptides that mimic the L2/HNK1  
CC epitope were isolated by screening phage peptide display libraries with  
CC antibodies to L2/HNK1. The present sequence is the consensus sequence  
CC for mimic peptides selected by HNK1 antibody. The peptides may be used  
CC for enhancing memory and for treating Alzheimer's disease or dementia.  
CC They may also be used for treating apoptosis, necrosis, Parkinson's  
CC disease, multiple sclerosis and acute and chronic spinal cord injury.  
XX  
SQ Sequence 6 AA:  
  
Query Match 54.5%; Score 18; DB 21; Length 6;  
Best Local Similarity 75.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 TRIF 4  
|:|:  
Db 1 trlf 4  
  
RESULT 10  
AAB13881  
ID AAB13881 standard; peptide; 6 AA.  
XX  
AC AAB13881;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE L2/HNK1 epitope mimic peptide homology sequence #1.  
XX  
KW L2/HNK1: epitope; nervous tissue; apoptosis; necrosis; antibody;  
KW Parkinson's disease; multiple sclerosis; spinal cord injury;  
XX Alzheimer's disease; dementia.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 5 /note= "Arg or nothing"  
XX  
PN WO200050447-A1.  
XX  
PD 31-AUG-2000.  
XX  
PF 24-FEB-2000; 2000MO-US04730.  
XX  
PR 24-FEB-1999; 99US-0121327.  
PR 24-FEB-1999; 99US-0256970.  
PR 23-SEP-1999; 99US-0155452.  
PR 23-SEP-1999; 99US-040431.  
PR 23-FEB-2000; 2000US-0511956.  
XX  
PA (ACOR-) ACORDA THERAPEUTICS.  
XX  
PI Schachner M, Neuberger TJ, Herzberg U, Simon M;  
XX WPI; 2000-579163/54.  
XX  
PT New isolated peptide which mimics a carbohydrate epitope is useful for  
PT neuroprotection  
XX  
PS Claim 8; Page 124; 214pp; English.  
XX  
CC The present invention relates to L2/HNK1 carbohydrate epitope mimic  
CC peptides. The L2/HNK1 epitope is predominantly expressed on glycolipids  
CC and glycoproteins from nervous tissue. Peptides that mimic the L2/HNK1  
CC epitope were isolated by screening phage peptide display libraries with  
CC antibodies to L2/HNK1. The present sequence is the homologous region  
CC between consensus sequences of mimic peptides isolated by different  
CC

CC antibodies. The peptides may be used for enhancing memory and for  
 CC treating Alzheimer's disease or dementia. They may also be used for  
 CC treating apoptosis, necrosis, Parkinson's disease, multiple sclerosis  
 CC and acute and chronic spinal cord injury.

XX Sequence 6 AA:

Query Match 54.5%; Score 18; DB 21; Length 6;  
 Best Local Similarity 75.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRIF 4  
 11:1  
 DB 1 trlf 4

#### RESULT 11

AA698766 standard; peptide; 7 AA.

XX AAC98766;

XX 21-SEP-2001 (first entry)

DE Human cell death protective cDNA clone CNI-00721 ORF19 peptide, SEQ:330.

XX Cell death protective; Apoptosis; necrosis; human; drug screening;  
 KW cell death-associated disorder; central nervous system disorder;  
 KW psychiatric disorder; neurological disorder; ischemia-related disorder;  
 KW stroke; cerebral infarction; ischemic encephalopathy;  
 KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;  
 KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;  
 KW vascular disease; ophthalmological disorder; diabetic retinopathy;  
 KW macular degeneration; hypertension; myocardial infarction;  
 KW atherosclerosis; respiratory disorder; asthma; transgenic animal;  
 KW chronic obstructive pulmonary disease; neoplastic condition; cancer;  
 KW benign tumor; anemia; gastrointestinal disorder; gastritis;  
 KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;  
 KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;  
 KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;  
 KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.

XX Homo sapiens.

XX WO200145638-A2.

XX 28-JUN-2001.

XX 11-DEC-2000; 2000WO-US35347.

XX 14-DEC-1999; 99US-0461697.

XX (COGE-) COCENT NEUROSCIENCE INC.

XX Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;

XX WPI; 2001-390297/41.

XX N-PSDB; AAH84281, AAH84300.

XX Novel protective sequence polynucleotides and polypeptides, used to  
 PT identify modulators of their expression and activity, which are used in  
 PT to treat central nervous system conditions, diseases and disorders -  
 XX  
 PS Claim 1; Fig 11S; 325pp; English.

XX Sequences AAH84132-AAH84370 represent human nucleic acid sequences which  
 CC protect against cell death (i.e., apoptosis or necrosis). Sequences  
 CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84226, AAH84265,  
 CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,  
 CC while the remaining nucleic acid sequences within the range given above  
 CC represent the open reading frames (ORFs) of these cDNA clones. Sequences  
 CC AA698610-AA698829 represent the polypeptides encoded by the cell death

CC protective ORFs. The cell death protective cDNA clones are able to  
 CC prevent, delay or reverse progression through the apoptotic or necrotic  
 CC pathways when injected into a cell predisposed to or undergoing cell  
 CC death. The cell death protective nucleic acids and polypeptides can be  
 CC used in the diagnosis and treatment of disorders associated with cell  
 CC death, and to screen for compounds which modulate their activity or  
 CC expression. Such modulators, preferably a small organic molecule, an  
 CC antibody, a ribozyme, or an antisense molecule, can also be used to treat  
 CC cell death-related diseases. Such diseases include those associated with  
 CC the central nervous system including psychiatric or neurological  
 CC disorders, especially ischemia-related conditions such as strokes, and  
 CC also includes neurodegenerative disorders such as Alzheimer's disease,  
 CC Huntington's disease, or Parkinson's disease. The modulators may also be  
 CC used to treat infections such as meningitis, malaria, or trypanosomiasis;  
 CC vascular diseases such as ischemic encephalopathy or cerebral  
 CC infarction; eye conditions such as diabetic retinopathy or macular  
 CC degeneration; hypertension; myocardial infarction; atherosclerosis;  
 CC respiratory conditions such as asthma or chronic obstructive pulmonary  
 CC disease; neoplastic conditions such as cancers or benign tumours; blood  
 CC cell conditions such as anaemia; gastrointestinal conditions such as  
 CC gastritis or ulcerative colitis; liver conditions such as biliary  
 CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;  
 CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's  
 CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune  
 CC system disorders such as acquired immunodeficiency syndrome (AIDS). The  
 CC nucleic acids may additionally be used to generate animal models of  
 CC cell death-associated disorders. The present sequence represents a  
 CC cell death protective polypeptide.

XX Sequence 7 AA:

Query Match 54.5%; Score 18; DB 22; Length 7;  
 Best Local Similarity 60.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 IFSKL 7  
 11:1  
 DB 2 IFSKL 6

XX

#### RESULT 12

AAW18503 standard; peptide; 5 AA.

XX AAW18503;

XX 19-FEB-1998 (first entry)

DE Amino-terminal peptide 5 associated with novel helicase.

XX Nucleic acid binding protein; helicase; leflunomide; assaying;  
 KW 4-(N-(4-trifluoromethylphenyl)-5-methyl-isoxazole; identification;  
 KW anticancer; antiatherosclerotic; immunosuppressant; sequencing;  
 KW antiinflammatory; antiviral; antifungal; antibacterial; treatment;  
 KW Alzheimer's disease; cancer; rheumatism; arthritis; determination;  
 KW atherosclerosis; osteoporosis; acute infection; chronic infection;  
 KW autoimmune disease; diabetes; organ transplant; isolation;  
 KW amino-terminal.

XX Homo sapiens.

XX DE19545126-A1.

XX 05-JUN-1997.

XX 04-DEC-1995; 95DE-1045126.

XX 04-DEC-1995; 95DE-1045126.

XX (FARH ) HOECHST AG.

XX Bartlett R, Kirschbaum B, Muellner S;

XX WPI: 1997-299388/28.  
 DR  
 XX  
 PT New nucleic acid binding protein with helicase activity - is  
 PT strongly induced by leflunomide, used to isolate specific binding  
 PT RNA and for identifying substances with anticancer, antiviral etc.  
 PT activities  
 XX  
 PS Example 4; Page 15; 28pp; German.  
 XX  
 CC The present sequence is an amino-terminal peptide associated with a  
 CC novel nucleic acid binding protein with helicase activity, the mRNA  
 CC of which (or its translation products) is strongly expressed in  
 CC presence of leflunomide, i.e.  
 CC 4-(N-(4-trifluoromethylphenyl)-5-methyl-isoxazole, or a compound  
 CC with similar activity. The helicase can be used in assay systems  
 CC to identify/discover anticancer, antitumor, antiviral, antifungal and  
 CC immunosuppressing, antiinflammatory, antiviral, cancer, cancer,  
 CC antibacterial agents, e.g. to treat Alzheimer's disease, cancer,  
 CC rheumatism, arthritis, atherosclerosis, osteoporosis,  
 CC acute/chronic infections, autoimmune disease, diabetes and  
 CC complications of organ transplants, and to isolate or determine the  
 CC sequences of specific binding RNA.  
 CC  
 XX  
 SO Sequence 5 AA:

Query Match 51.5%; Score 17; DB 18; Length 5;  
 Best Local Similarity 60.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 RIFSK 6  
 :|:|  
 Db 1 klfqk 5

## RESULT 13

AAW68421 standard; peptide; 6 AA.

AC AAW68421;

DT 23-NOV-1998 (first entry)

DE Peptide #3 from cyclic screening method for physiological activity.

XX Screening method; physiological activity; hexapeptide; algorithm;

KW crossover; mutation; enzyme inhibitor; drug; foodstuff.

OS Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note- "acylated N-terminus"

FT Modified-site 6 /note- "amidated C-terminus"

XX WO9804580-A1.

XX 05-FEB-1998.

XX 23-JUL-1997; 97WO-JP02535.

XX 26-JUL-1996; 96JP-0198096.

XX (YAMA ) YAMANOUCHI PHARM CO LTD.

XX Karube M, Yokobayashi Y;

XX WPI: 1998-130616/12.

PT Method for screening peptide(s) with high physiological activity -  
 PT by determining activity, dropping least active, changing sequence in

PT remaining peptide(s) and cycling through this sequence to obtain  
 PT peptide(s) with high activity  
 XX  
 PS Disclosure; Page 9; 19pp; Japanese.

XX Peptides AAW68419-W68434 represent peptides obtained by the cyclic  
 CC screening method of the invention. The method comprises determining  
 CC the level of a desired physiological activity in a series of  
 CC hexapeptides and selecting those with the highest activities. The  
 CC peptides are then processed using a genetic algorithm computer program  
 CC which: (i) exchanges one or more amino acid residues between pairs of  
 CC peptides (crossover) and/or (ii) substitutes amino acid residues in a  
 CC frequency of about 3% of the total number of residues. The altered  
 CC peptides are synthesized and the screening cycle is repeated until  
 CC peptides with the desired level of activity are generated. Peptides with  
 CC a high activity (e.g. as enzyme inhibitors) can be used to produce  
 CC compositions for use in the drug, foodstuff and other industries.  
 CC  
 XX  
 SO Sequence 6 AA:

Query Match 51.5%; Score 17; DB 19; Length 6;  
 Best Local Similarity 75.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIF 4  
 :|:|  
 Db 1 klf 4

## RESULT 14

AAW68428 standard; peptide; 6 AA.

AC AAW68428;

DT 23-NOV-1998 (first entry)

DE Peptide #10 from cyclic screening method for physiological activity.

XX Screening method; physiological activity; hexapeptide; algorithm;

KW crossover; mutation; enzyme inhibitor; drug; foodstuff.

OS Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note- "acylated N-terminus"

FT Modified-site 6 /note- "amidated C-terminus"

XX WO9804580-A1.

XX 05-FEB-1998.

XX 23-JUL-1997; 97WO-JP02535.

XX 26-JUL-1996; 96JP-0198096.

XX (YAMA ) YAMANOUCHI PHARM CO LTD.

XX Karube M, Yokobayashi Y;

XX WPI: 1998-130616/12.

PT Method for screening peptide(s) with high physiological activity -  
 PT by determining activity, dropping least active, changing sequence in  
 PT remaining peptide(s) and cycling through this sequence to obtain  
 PT peptide(s) with high activity

PS Disclosure; Page 9; 19pp; Japanese.

CC Peptides AAW68419-W68434 represent peptides obtained by the cyclic  
 CC screening method of the invention. The method comprises determining  
 CC the level of a desired physiological activity in a series of  
 CC hexapeptides and selecting those with the highest activities. The  
 CC peptides are then processed using a genetic algorithm computer program  
 CC which: (i) exchanges one or more amino acid residues between pairs of  
 CC peptides (crossover) and/or (ii) substitutes amino acid residues in a  
 CC peptide for different ones (mutation) (preferably with a mutation  
 CC frequency of about 3% of the total number of residues). The altered  
 CC peptides are synthesised and the screening cycle is repeated until  
 CC peptides with the desired level of activity are generated. Peptides with  
 CC a high activity (e.g. as enzyme inhibitors) can be used to produce  
 CC compositions for use in the drug, foodstuff and other industries.  
 XX  
 SQ Sequence 6 AA;

Query Match 51.5%; Score 17; DB 19; Length 6;  
 Best Local Similarity 75.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRIF 4  
 1:11  
 Db 2 TKIF 5

## RESULT 15

AA17746  
 ID AA17746 standard; peptide: 6 AA.  
 AC AA17746;

DT 10-AUG-1999 (first entry)  
 XX

DE Calreticulin binding peptide SEQ ID NO:65.  
 XX

KM Calreticulin; binding; hormone responsiveness; modulation; mimetic;  
 KW inhibitor; gene therapy; cancer; osteoporosis; pharmaceutical;  
 KM chronic inflammatory disease; promyelocytic leukaemia; arthritis;  
 KM bone disease; hormone receptor binding.  
 XX

OS Synthetic.  
 XX

PN WO9905172-A2.  
 XX

PD 04-FEB-1999.  
 XX

PF 24-JUL-1998; 98WO-CA00715.  
 XX

PR 24-JUL-1997; 97US-0900241.  
 XX

PA (DEDH/) DEDHAR S.  
 PA (DOER/) DOERSEN C W.  
 PA (MAZU/) MAZUR A W.  
 XX

PI Dedhar S, Doersen CW, Mazur AW;  
 XX

DR WPI: 1999-142854/12.  
 XX

PT New peptides that bind calreticulin and modulate gene expression -  
 PT are activated by hormone receptors, useful in the treatment of  
 PT cancer, chronic inflammation and osteoporosis  
 XX

PS Example 15; Page 44; 64pp; English.  
 XX

CC The present invention describes peptides (I) that bind to calreticulin  
 CC (CR), specifically comprising the sequence: KGX1X2X3R, where one or more  
 CC X = basic amino acid (aa). (I) are used to treat cancer (particularly of  
 CC the prostate or breast, or promyelocytic leukaemia), chronic  
 CC inflammation (e.g. arthritis) or osteoporosis, also they can be used to  
 CC treat other bone diseases. (I) act by modulating binding of hormone  
 CC receptors (HR) to DNA. The present sequence represents a peptide used  
 CC in the exemplification of the present invention.

XX  
 SQ Sequence 6 AA;

Query Match 51.5%; Score 17; DB 20; Length 6;  
 Best Local Similarity 40.0%; Pred. No. 6.4e+05;  
 Matches 2; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 RIFSK 6  
 1:11  
 Db 1 KVFAK 5

Search completed: July 15, 2002, 13:25:25  
 Job time: 1456 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:26:12 ; Search time 40.89 Seconds  
(without alignments)  
4.181 Million cell updates/sec

Title: US-09-712-819A-2

Sequence: 1 TRIFSKL 7

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 39160

Minimum DB seq length: 0  
Maximum DB seq length: 7

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2-6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2-6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2-6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2-6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2-6/ptodata/2/1aa/PCUT6.COMB.pep.\*  
6: /cgn2-6/ptodata/2/1aa/Backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	69.7	7	2	US-09-070-756-9
2	19	57.6	6	1	US-07-634-641-13
3	18	54.5	6	4	US-08-540-922D-6
4	18	54.5	7	4	US-09-661-697-330
5	17	51.5	5	1	US-07-858-222A-5
6	17	51.5	5	2	US-08-760-075A-5
7	17	51.5	5	4	US-09-338-346-5
8	17	51.5	6	1	US-07-943-709-117
9	17	51.5	6	1	US-07-943-709-118
10	17	51.5	7	1	US-08-253-854-7
11	17	51.5	7	2	US-08-392-973A-35
12	17	51.5	7	4	US-09-046-373-10
13	16	48.5	4	1	US-08-419-903A-2
14	16	48.5	4	3	US-09-020-299-4
15	16	48.5	5	1	US-08-118-135A-3
16	16	48.5	5	2	US-08-392-973A-10
17	16	48.5	5	4	US-09-177-249-266
18	16	48.5	6	2	US-08-392-973A-17
19	16	48.5	6	2	US-08-392-973A-18
20	16	48.5	6	4	US-09-106-216-57
21	16	48.5	7	1	US-08-050-232-9
22	16	48.5	7	1	US-08-457-274A-13
23	16	48.5	7	2	US-08-661-767-9
24	16	48.5	7	2	US-08-392-973A-23
25	16	48.5	7	2	US-08-392-973A-24
26	16	48.5	7	2	US-08-392-973A-25
27	16	48.5	7	4	US-09-031-902-10

28	16	48.5	7	5	PCT-US95-05758-13	Sequence 13, Appl
29	15	45.5	4	1	US-08-450-384-21	Sequence 21, Appl
30	15	45.5	4	6	517513-1	Patent No. 517513
31	15	45.5	5	1	US-08-200-900A-33	Sequence 9, Appl
32	15	45.5	5	2	US-08-392-973A-9	Sequence 11, Appl
33	15	45.5	5	2	US-08-896-605A-3	Sequence 3, Appl
34	15	45.5	5	2	US-08-896-501A-3	Sequence 3, Appl
35	15	45.5	5	2	US-08-483-506A-6	Sequence 3, Appl
36	15	45.5	5	2	PCT-US94-00616-33	Sequence 33, Appl
37	15	45.5	6	1	US-07-718-577-13	Sequence 13, Appl
38	15	45.5	6	1	US-08-188-228-3	Sequence 3, Appl
39	15	45.5	6	1	US-08-332-643-3	Sequence 3, Appl
40	15	45.5	6	1	US-08-332-638-3	Sequence 3, Appl
41	15	45.5	6	1	US-08-424-957-6	Sequence 6, Appl
42	15	45.5	6	1	US-08-392-973A-16	Sequence 16, Appl
43	15	45.5	6	2	US-08-392-973A-19	Sequence 19, Appl
44	15	45.5	6	2	US-07-662-764D-23	Sequence 23, Appl
45	15	45.5	6	2		

#### ALIGNMENTS

RESULT 1  
US-09-070-756-9  
Sequence 9, Application US/09070756  
Patent No. 5917012  
GENERAL INFORMATION:  
APPLICANT: Nishikata, Makoto  
TITLE OF INVENTION: NOVEL PEPTIDE DERIVATIVES  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: DAVID S. RESNICK; DIKE, BRONSTEIN, ROBERTS &  
ADDRESS: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/070,756  
FILING DATE: 30 April 1998  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Resnick, David S.  
REGISTRATION NUMBER: 34,235  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-070-756-9  
Query Match 69.7%; Score 23; DB 2; Length 7;  
Best Local Similarity 80.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 RFSK 6  
1:111  
Db 2 RFSK 6  
RESULT 2

US-07-634-641-13  
Sequence 13, Application US/07634641  
Patent No. 5386011  
GENERAL INFORMATION:  
APPLICANT: Wiedeman, Paul E.  
APPLICANT: Kawai, Megumi  
APPLICANT: Lully, Jay R.  
APPLICANT: Of, Yat-Sun  
APPLICANT: Wagner, Rolf  
TITLE OF INVENTION: Hexa- and Heptapeptide Anaphylatoxin  
TITLE OF INVENTION: Receptor Ligands  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: One Abbott Park Road  
CITY: No. 5386011th Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/634,641  
FILING DATE: 19901227  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Janssen, Jerry F.  
REGISTRATION NUMBER: 29,175  
REFERENCE/DOCKET NUMBER: 4934.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 937-4558  
TELEFAX: (708) 937-9556  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptidic  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 2  
OTHER INFORMATION: /note="XAA at position 2 is an  
OTHER INFORMATION: L-cyclohexylalanyl residue"  
LOCATION: 4  
OTHER INFORMATION: /note="XAA at position 4 is an  
OTHER INFORMATION: L-cyclohexylalanyl residue"  
US-07-634-641-13

Query Match 57.6%; Score 19; DB 1; Length 6;  
Best Local Similarity 66.7%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 RIFSKL 7  
DB 1 RKFEXL 6

RESULT 3  
US-08-540-922D-6  
Sequence 6, Application US/08540922D  
Patent No. 6284476  
GENERAL INFORMATION:  
APPLICANT: Boon-Falleur, Thierry, Brichard, Vincent; Van  
APPLICANT: Pel, Aline; De Jaen, Etienne; Couille, Pierre;  
APPLICANT: Renaud Jean-Christophe; Wolter, Thomas; and  
APPLICANT: Lethe, Bernard.

TITLE OF INVENTION: METHOD OF IDENTIFYING INDIVIDUALS SUFFERING  
FROM A CELLULAR ABNORMALITY SOME OF WHOSE  
TITLE OF INVENTION: ABNORMAL CELLS PRESENT COMPLEXES OF HUMAN  
LEUCOCYTE ANTIGEN TYROSINASE DERIVED  
PEPTIDES, AND METHODS FOR TREATING SAID  
INDIVIDUALS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felle & Lynch  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch 1.44 MB storage diskette  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/540,922D  
FILING DATE: October 11, 1995  
CLASSIFICATION: 514  
APPLICATION NUMBER: 08/054,714  
FILING DATE: 28 April 1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 07/994,928  
FILING DATE: 22 December 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary Anne Schofield  
REGISTRATION NUMBER: 36,669  
REFERENCE/DOCKET NUMBER: LUD 5299.5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6  
TYPE: amino acids  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: Protein  
FEATURE:  
OTHER INFORMATION: SEQ of aa corresponding to nt 1714-1731  
OTHER INFORMATION: nt of SEQ ID NO: 1  
US-08-540-922D-6

Query Match 54.5%; Score 18; DB 4; Length 6;  
Best Local Similarity 66.7%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 RIFSKL 7  
DB 1 RPFAXL 6

RESULT 4  
US-09-461-697-330  
Sequence 330, Application US/09461697  
Patent No. 6277974  
GENERAL INFORMATION:  
APPLICANT: COGENT NEUROSCIENCE, INC.  
APPLICANT: Lo, Donald C.  
APPLICANT: Barney, Shawn  
APPLICANT: Thomas, Mary Beth  
APPLICANT: Portbury, Stuart D.  
APPLICANT: Putnam, Kasuri  
APPLICANT: Katz, Lawrence C.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING



TITLE OF INVENTION: CELL DEATH  
FILE REFERENCE: 10001-005-999  
CURRENT APPLICATION NUMBER: US/09/461,697  
CURRENT FILING DATE: 1999-12-14  
NUMBER OF SEQ ID NOS: 466  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 330  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-461-697-330

Query Match 54.5%; Score 18; DB 4; Length 7;  
Best Local Similarity 60.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 IFSKL 7  
DB 2 LFSRL 6

RESULT 5  
US-07-958-222A-5  
Sequence 5, Application US/07958222A  
Patent No. 5332668  
GENERAL INFORMATION:  
APPLICANT: YAMASHITA, TAKASHI  
APPLICANT: HIGASHI, SUSUMU  
APPLICANT: HIGASHI, TOSHIO  
APPLICANT: MACHIDA, HARUO  
APPLICANT: IMASAKI, SHINJIRO  
APPLICANT: BEPPU, TERUHIRO  
TITLE OF INVENTION: PROTEASE WITH LOW THERMOSTABILITY AND RELATED  
TITLE OF INVENTION: PRODUCT THEREOF AND METHOD FOR PRODUCING THE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: KNOBBE, MARTENS, OLSON AND BEAR  
STREET: 620 NEWPORT CENTER DRIVE 16TH FLOOR  
CITY: NEWPORT BEACH  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/958,222A  
FILING DATE: 19921008  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 3-263878  
FILING DATE: 11-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E.  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: SATOT14.001AUS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX: 183513 KNOBBE NPBH  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
HYPOTHEICAL: NO  
US-07-958-222A-5

Query Match 51.5%; Score 17; DB 1; Length 5;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 IFSK 6  
DB 2 LFSK 5

RESULT 6  
US-08-760-075A-5

Sequence 5, Application US/08760075A  
Patent No. 5942429

GENERAL INFORMATION:

APPLICANT: KIRSCHBAUM, Bernd

APPLICANT: MUELLNER, Stefan

APPLICANT: BARTLETT, Robert

TITLE OF INVENTION: NOVEL DEAH-BOX PROTEINS

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/760,075A

FILING DATE: 04-DEC-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE 19545126.0

FILING DATE: 04-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: GRANADOS, Patricia D.

REGISTRATION NUMBER: 33,683

REFERENCE/DOCKET NUMBER: 18748/309

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: 904136

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 5 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-760-075A-5

Query Match 51.5%; Score 17; DB 2; Length 5;  
Best Local Similarity 60.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIFSK 6  
DB 1 KIFOK 5

RESULT 7  
US-09-338-546-5

Sequence 5, Application US/09338546  
Patent No. 6251645

GENERAL INFORMATION:

APPLICANT: KIRSCHBAUM, Bernd

APPLICANT: MUELLNER, Stefan

APPLICANT: BARTLETT, Robert

;; TITLE OF INVENTION: NOVEL DEAH-BOX PROTEINS  
;; NUMBER OF SEQUENCES: 38  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Foley & Lardner  
;; STREET: 3000 K Street, N.W., Suite 500  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20007-5109  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/338,546  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/760,075  
;; FILING DATE: 04-DEC-1996  
;; APPLICATION NUMBER: DE 19545126.0  
;; FILING DATE: 04-DEC-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: GRANDOS, Patricia D.  
;; REGISTRATION NUMBER: 33,683  
;; REFERENCE/DOCKET NUMBER: 18/48/309  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202)672-5300  
;; TELEFAX: (202)672-5399  
;; TELEX: 904136  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 5 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-09-338-546-5

Query Match 51.5%; Score 17; DB 4; Length 5;  
Best Local Similarity 60.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RFSK 6  
;|||  
Db 1 KIRFX 5

RESULT 8  
US-07-943-709-117  
; Sequence 117, Application US/07943709  
; Patent No. 5556762  
; GENERAL INFORMATION:  
; APPLICANT: Pinilla, Clemencia  
; APPLICANT: Appel Jr., Jon R.  
; APPLICANT: Blondelle, Sylvie  
; APPLICANT: Dooley, Colette T.  
; APPLICANT: Eichler, Jutta  
; APPLICANT: Houghten, Richard A.  
; TITLE OF INVENTION: SCANNING SYNTHETIC PEPTIDE COMBINATORIAL  
; LIBRARIES: OLIGOPEPTIDE MIXTURE SETS HAVING ONE  
; TITLE OF INVENTION: PREDETERMINED RESIDUE AT A SINGLE, PREDETERMINED  
; POSITION, PREDETERMINED POSITION,  
; NUMBER OF SEQUENCES: 119  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dressler, Goldsmith, Shore, Suter &  
; STREET: 180 No. 5556762th Stetson, Suite 4700  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60601

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/943,709  
;; FILING DATE: 19920911  
;; CLASSIFICATION: 530  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/797,551  
;; FILING DATE: 19-NOV-1991  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Gansson, Edward P.  
;; REGISTRATION NUMBER: 29,381  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (312) 616-5400  
;; TELEFAX: (312) 616-5460  
;; INFORMATION FOR SEQ ID NO: 117:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 6 amino acids  
;; TYPE: AMINO ACID  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 1  
;; OTHER INFORMATION: /note= "Xaa is Ac-Met."  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 6  
;; OTHER INFORMATION: /note= "Xaa is Thr-NH2."  
;; US-07-943-709-117

Query Match 51.5%; Score 17; DB 1; Length 6;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRIF 4  
;|||  
Db 2 TKIF 5

RESULT 9  
US-07-943-709-118  
; Sequence 118 Application US/07943709  
; Patent No. 5556762  
; GENERAL INFORMATION:  
; APPLICANT: Pinilla, Clemencia  
; APPLICANT: Appel Jr., Jon R.  
; APPLICANT: Blondelle, Sylvie  
; APPLICANT: Dooley, Colette T.  
; APPLICANT: Eichler, Jutta  
; APPLICANT: Houghten, Richard A.  
; TITLE OF INVENTION: SCANNING SYNTHETIC PEPTIDE COMBINATORIAL  
; LIBRARIES: OLIGOPEPTIDE MIXTURE SETS HAVING ONE  
; TITLE OF INVENTION: PREDETERMINED RESIDUE AT A SINGLE, PREDETERMINED  
; POSITION, PREDETERMINED POSITION,  
; NUMBER OF SEQUENCES: 119  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dressler, Goldsmith, Shore, Suter &  
; STREET: 180 No. 5556762th Stetson, Suite 4700  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/943,709  
FILING DATE: 19920911  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/797,551  
FILING DATE: 19-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Gamson, Edward P.  
REGISTRATION NUMBER: 29,381  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 616-5400  
TELEFAX: (312) 616-5460  
INFORMATION FOR SEQ ID NO: 118:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /note="Xaa is Ac-Thr."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 6  
OTHER INFORMATION: /note="Xaa is Thr-NH2."  
US-07-943-709-118

Query Match 51.5%; Score 17; DB 1; Length 6;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRIF 4  
1:11  
DB 2 TRIF 5

RESULT 10  
US-08-253-854-7  
Sequence 7; Application US/08253854  
Patent No. 5504190  
GENERAL INFORMATION:  
APPLICANT: Houghten, Richard A.  
APPLICANT: Cuervo, Julio H.  
APPLICANT: Pinilla, Clemencia  
APPLICANT: Appel Jr., Jon R.  
APPLICANT: Blondelle, Sylvie  
TITLE OF INVENTION: Synthesis of Equimolar Multiple  
NUMBER OF SEQUENCES: 91  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dressler, Goldsmith, Shore, Sulkner &  
ADDRESSEE: Milmanow, Ltd.  
STREET: 180 No. 5504190th Stetson Avenue, Suite 4700  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/253,854  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gamson, Edward P.  
REGISTRATION NUMBER: 29,381  
REFERENCE/DOCKET NUMBER: PRL.0003

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 616-5400  
TELEFAX: (312) 616-5460  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-253-854-7

Query Match 51.5%; Score 17; DB 1; Length 7;  
Best Local Similarity 60.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TRIFS 5  
1:11  
DB 3 TRIFS 7

RESULT 11  
US-08-392-973A-35  
Sequence 35; Application US/08392973A  
Patent No. 5843453  
GENERAL INFORMATION:  
APPLICANT: Holder, Andrew T.  
APPLICANT: HEATTE, James  
TITLE OF INVENTION: GROWTH HORMONE POTENTIATING MOLECULES  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HALE AND DORR LLP  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: United States of America  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/392,973A  
FILING DATE: 25-APR-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: BAKER, Hollis L.  
REGISTRATION NUMBER: 31,321  
REFERENCE/DOCKET NUMBER: 102286.301  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-526-6000  
TELEFAX: 617-526-5000  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..7  
OTHER INFORMATION: /note="94-100 REGION OF RAT  
GROWTH HORMONE"  
US-08-392-973A-35

Query Match 51.5%; Score 17; DB 2; Length 7;  
Best Local Similarity 60.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIPS 5  
DB 1 SRIPT 5

## RESULT 12

US-09-046-373-10  
; Sequence 10, Application US/09046373  
; Patent No. 6235714  
; GENERAL INFORMATION:  
; APPLICANT: Sudhir Paul  
; APPLICANT: Larry J. Smith  
; APPLICANT: Gennady Golodnov  
; TITLE OF INVENTION: Methods for Identifying Inducers and  
; TITLE OF INVENTION: Inhibitors of Catalytic Antibodies, Compositions and Their  
; FILE REFERENCE: UNMC 63123  
; CURRENT APPLICATION NUMBER: US/09/046.373  
; CURRENT FILING DATE: 1998-03-23  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 10  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-046-373-10

Query Match 51.5%; Score 17; DB 4; Length 7;  
Best Local Similarity 40.0%; Pred. No. 1.7e+05;  
Matches 2; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 3 IFSKL 7  
DB 2 VFNKI 6

## RESULT 13

US-08-419-903A-2  
; Sequence 2, Application US/08419903A  
; Patent No. 5753226  
; GENERAL INFORMATION:  
; APPLICANT: Greene, Mark I.  
; APPLICANT: Colasarellis, George  
; TITLE OF INVENTION: METHODS OF ENHANCING EPITHELIAL CELL  
; TITLE OF INVENTION: PROLIFERATION  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5753226rls  
; STREET: One Liberty Place 46th Floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: United States of America  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/419.903A  
; FILING DATE: 11-APR-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Deluca, Mark  
; REGISTRATION NUMBER: 33,229  
; REFERENCE/DOCKET NUMBER: UPN-1705  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-568-3100  
; TELEFAX: 215-568-3439  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids

; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-419-903A-2

Query Match 48.5%; Score 16; DB 1; Length 4;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 ESKL 7  
DB 1 FNKL 4

## RESULT 14

US-09-020-299-4  
; Sequence 4, Application US/09020299  
; Patent No. 6090381  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG, Shui-on  
; APPLICANT: OU, Zhengxing  
; TITLE OF INVENTION: STIMULATION OF AN IMMUNE RESPONSE WITH ANTIBODIES  
; FILE REFERENCE: 018733/0814  
; CURRENT APPLICATION NUMBER: US/09/020.299  
; CURRENT FILING DATE: 1998-02-06  
; EARLIER APPLICATION NUMBER: US 60/037,908  
; EARLIER FILING DATE: 1997-02-11  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-020-299-4

Query Match 48.5%; Score 16; DB 3; Length 4;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIF 4  
DB 1 SRIPT 4

## RESULT 15

US-08-118-135A-3  
; Sequence 3, Application US/08118135A  
; Patent No. 5437982  
; GENERAL INFORMATION:  
; APPLICANT: Carterall, W.A. and Esholtz, G.  
; TITLE OF INVENTION: Specific Inactivation Gate Inhibitors of the Sodium Channel  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101-2347  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette-5.25 inch, 1.2mb storage  
; COMPUTER: IBM PC/386 Compatible  
; OPERATING SYSTEM: MS-DOS 4.01  
; SOFTWARE: Word 5.5-1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/118.135A  
; FILING DATE: September 8, 1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA: No. 5437982e  
; ATTORNEY/AGENT INFORMATION:

NAME: Broderick, Thomas F.  
 REGISTRATION NUMBER: 31,332  
 REFERENCE/DOCKET NUMBER: UOFW16993  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 1-206-682-8100; 1-206-224-0709 (direct)  
 TELEFAX: 1-206-224-0779  
 TELEX: 4938023  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 DESCRIPTION: page 5, line 4; KIFWK  
 US-08-118-135A-3

Query Match 48.5%; Score 16; DB 1; Length 5;  
 Best Local Similarity 60.0%; Pred. No. 1.7e+05;  
 Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Oy 2 RIFSK 6  
 :|||  
 Db 1 KIFWK 5

Search completed: July 15, 2002, 13:26:13  
 Job time: 493 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 15, 2002, 13:27:08 ; Search time 49.13 Seconds

(without alignments)  
13,691 Million cell updates/sec

Title: US-09-712-819A-3

Perfect score: 35

Sequence: 1 FYOLAVL 7

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 28138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 455

Minimum DB seq length: 0

Maximum DB seq length: 7

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	37.1	3	3	669328
2	13	37.1	5	2	B44823
3	13	37.1	7	2	E30608
4	13	37.1	7	2	E33932
5	12	34.3	5	2	I40469
6	12	34.3	7	2	B39127
7	11	31.4	6	2	A37765
8	11	31.4	7	2	S18630
9	10	28.6	4	2	S18401
10	10	28.6	4	2	A32039
11	10	28.6	4	2	T01273
12	10	28.6	6	2	S60293
13	10	28.6	6	2	B35640
14	10	28.6	6	2	I49421
15	10	28.6	6	2	A43129
16	10	28.6	7	2	S55348
17	10	28.6	7	2	PN0150
18	10	28.6	7	2	PS0254
19	10	28.6	7	2	E48394
20	10	28.6	7	2	B48394
21	9	25.7	4	2	I38888
22	9	25.7	5	2	A60521
23	9	25.7	5	2	S68326
24	9	25.7	6	2	I48126
25	9	25.7	7	2	S33246
26	9	25.7	7	2	A38081
27	8	22.9	3	3	A22565
28	8	22.9	4	2	A37832
29	8	22.9	4	2	PT0697

## ALIGNMENTS

30	8	22.9	5	2	JN0860	peptidyl-dipeptid
31	8	22.9	5	2	E60274	major protein anti
32	8	22.9	5	2	P22565	R-phycocerythrin ga
33	8	22.9	5	2	PS0324	ribulose-bisphosph
34	8	22.9	5	2	S55237	zinc-binding prote
35	8	22.9	5	2	JH0253	gut pentapeptide -
36	8	22.9	5	2	PT0590	T-cell receptor be
37	8	22.9	5	2	B44817	34.5K structural p
38	8	22.9	5	2	D44817	hydrogensulfite re
39	8	22.9	6	2	S11556	28K ubiquitin-immu
40	8	22.9	6	2	A43766	proctamine P1 - gor
41	8	22.9	6	2	I37027	19 mu chain D regl
42	8	22.9	6	2	B33932	T-cell receptor ga
43	8	22.9	7	1	A41946	hypothalamic hepta
44	8	22.9	7	1	NYP67	dermorphin (Trp-4,
45	8	22.9	7	2	S21230	

RESULT 1  
S68328  
blood cell protein A - Molgula manhatensis (fragment)  
C:Species: Molgula manhatensis  
C:Date: 15-Jun-2001 #sequence\_revision 15-Jun-2001 #text\_change 15-Jun-2001  
C:Accession: S68328  
R:Taylor, S.W.; Ross, M.M.; Waite, J.H.  
Arch. Biochem. Biophys. 324, 228-240, 1995  
A:Title: Novel 3,4-di- and 3,4,5-trihydroxyphenylalanine-containing polypeptides from  
A:Reference number: S68325; MUID:96132650  
A:Accession: S68328  
A:Molecule type: protein  
A:Residues: 1-3 <TAV>

Query Match 37.1%; Score 13; DB 3; Length 3;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FY 2  
DB 2 FY 3

RESULT 2  
B44823  
synaptosomal-associated protein SNAP-25 peptide 10A - rabbit (fragment)  
N:Alternate names: superprotein peptide 10A  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 31-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 15-Jun-1996  
C:Accession: B44823  
R:Rowley, A.; Liu, W.S.; Baillinger, C.; Willard, M.B.  
J. Neurosci. 11, 3412-3421, 1991  
A:Title: The major 35S-methionine-labeled rapidly transported protein (superprotein)  
A:Reference number: A44823; MUID:92044785  
A:Accession: B44823  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-5 <LOE>  
A:Experimental source: visual tissue  
A:Note: sequence extracted from NCBI backbone (NCBIP:64255)  
C:Keywords: membrane trafficking

Query Match 37.1%; Score 13; DB 2; Length 5;  
Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 OLAV 6  
DB 2 QMAI 5

RESULT 3  
E30608  
Ig kappa chain V-II region (Gag) - human (fragment)

C:Species: Homo sapiens (man)  
C>Date: 29-Jun-1989 #sequence\_revision 29-Jun-1989 #text\_change 16-Aug-1996  
C:Accession: E30608  
R:Gonl, F.R.; Chen, P.P.; McGinnis, D.; Arjona, M.L.; Fernandez, J.; Carson, D.; Sold  
J. Immunol. 142, 3158-3163, 1989  
A:Title: Structural and Idiotypic characterization of the L chains of human IgM autocal  
A:Reference number: A30601; MUID:89215279  
A:Accession: E30608  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-7 <GON>  
C:Keywords: heterotetramer; immunoglobulin

Query Match 37.1%; Score 13; DB 2; Length 7;  
Best Local Similarity 40.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 QIALT 7  
Db 1 EIVLT 5

RESULT 4  
E33932  
Ig mu chain D region (E7) - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 09-Mar-1990 #sequence\_revision 09-Mar-1990 #text\_change 16-Aug-1996  
C:Accession: E33932  
R:Baccala, R.; Vo Quang, T.; Gilbert, M.; Ternynck, T.; Avrameas, S.  
Proc. Natl. Acad. Sci. U.S.A. 86, 4624-4628, 1989  
A:Title: Two murine natural polyreactive autoantibodies are encoded by nonmutated germ-  
A:Reference number: A33932; MUID:89282823  
A:Accession: E33932  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-7 <BAC>  
A:Cross-references: GB:M27106  
C:Keywords: immunoglobulin

Query Match 37.1%; Score 13; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FY 2  
Db 3 FY 4

RESULT 5  
I40469  
dnaXX-like protein - Bacillus subtilis (fragment)  
C:Species: Bacillus subtilis  
C>Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 21-Jul-2000  
C:Accession: I40469  
R:Struck, U.C.; Hartmann, R.K.; Toschka, H.Y.; Erdmann, V.A.  
Mol. Gen. Genet. 215, 478-482, 1989  
A:Title: Transcription and processing of Bacillus subtilis small cytoplasmic RNA.  
A:Reference number: I40469; MUID:89218958  
A:Accession: I40469  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5 <RES>  
A:Cross-references: EMBL:X14796; NID:g40130; PIDD:CAA32902.1; PID:g4376204  
C:Genetic: 1  
A:Start codon: CTC

Query Match 34.3%; Score 12; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 YQ 3  
Db 3 YQ 4

RESULT 6  
B39127  
phosphotransferase system enzyme II (EC 2.7.1.69) - Escherichia coli (fragment)  
C:Species: Escherichia coli  
C>Date: 27-Nov-1991 #sequence\_revision 27-Nov-1991 #text\_change 08-Oct-1999  
C:Accession: B39127  
R:Hardesty, C.; Ferran, C.; DiRienzo, J.M.  
J. Bacteriol. 173, 449-456, 1991  
A:Title: Plasmid-mediated sucrose metabolism in Escherichia coli: characterization of  
A:Reference number: A39127; MUID:91100329  
A:Accession: B39127  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-7 <BAR>  
A:Cross-references: GB:M38416; NID:g155142; PIDD:AAA98418.1; PID:g155144  
C:Keywords: phosphotransferase

Query Match 34.3%; Score 12; DB 2; Length 7;  
Best Local Similarity 40.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYOLA 5  
Db 3 FEOIS 7

RESULT 7  
A37765  
hypothetical protein (csma 5' region) - Chloroflexus aurantiacus (fragment)  
C:Species: Chloroflexus aurantiacus  
C>Date: 31-May-1991 #sequence\_revision 31-May-1991 #text\_change 30-Sep-1993  
C:Accession: A37765  
R:Theroux, S.J.; Redlinger, T.E.; Fuller, R.C.; Robinson, S.J.  
J. Bacteriol. 172, 4497-4504, 1990  
A:Title: Gene encoding the 5.7-kilodalton chlorosome protein of Chloroflexus aurantia  
A:Reference number: A37765; MUID:90330558  
A:Accession: A37765  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-6 <THE>  
A:Cross-references: GB:M33964

Query Match 31.4%; Score 11; DB 2; Length 5;  
Best Local Similarity 66.7%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 QLA 5  
Db 2 QNA 4

RESULT 8  
S19630  
ribosomal protein L30 - Streptomyces griseus (fragment)  
C:Species: Streptomyces griseus  
C>Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 06-Jun-1997  
C:Accession: S19630  
R:Oehl, K.  
Int. J. Syst. Bacteriol. 42, 144-150, 1992  
A:Title: Electrophoretic heterogeneity of ribosomal protein AT-L30 among actinomycete  
A:Reference number: S19630; MUID:92144363



A:Accession: S19630  
 A:Molecule type: protein  
 A:Residues: 1-7 <OCH>  
 A:Experimental source: Strain IFO 13189  
 C:Superfamily: Escherichia coli ribosomal protein L30  
 C:Keywords: protein biosynthesis; ribosome

Query Match 31.4%; Score 11; DB 2; Length 7;  
 Best Local Similarity 40.0%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 QALMT 7  
 DB 2 RLKIT 6

RESULT 9  
 S18401  
 C:Species: Canis lupus familiaris (dog)  
 C:Date: 19-Mar-1997 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
 C:Accession: S18401  
 R:Donor: A.; Vassart, G.; Christophe, P.  
 Biochim. Biophys. Acta 1090, 235-237, 1991  
 A:Title: Isolation and characterization of the canine thyroglobulin gene promoter region  
 A:Reference number: S18401; MUID:92031697  
 A:Accession: S18401  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-4 <DON>  
 A:Cross-references: GB:S61184; NID:9237714; PIDN:AB20127.1; PID:9237715  
 C:Superfamily: thyroglobulin; cholinesterase homology; thyroglobulin type I repeat homol  
 C:Keywords: duplication; iodine; thyroid gland; thyroid hormone biosynthesis

Query Match 28.6%; Score 10; DB 2; Length 4;  
 Best Local Similarity 66.7%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 LAL 6  
 DB 1 MAL 3

RESULT 10  
 A32039  
 C:Species: melanocyte-stimulating hormone release-inhibiting factor 1 - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 31-Jul-1989 #sequence\_revision 31-Jul-1989 #text\_change 18-Aug-2000  
 C:Accession: A32039  
 R:Horvath, A.; Kastin, A.J.  
 J. Biol. Chem. 264, 2173-2179, 1989  
 A:Title: Isolation of tyrosine-melanocyte-stimulating hormone release-inhibiting factor  
 A:Reference number: A32039; MUID:89123285  
 A:Accession: A32039  
 A:Molecule type: protein  
 A:Residues: 1-4 <HOR>  
 A:Experimental source: brain  
 C:Superfamily: unassigned animal peptides  
 C:Keywords: amidated carboxyl end  
 F:4/Modified site: amidated carboxyl end (gly) #status experimental

Query Match 28.6%; Score 10; DB 2; Length 4;  
 Best Local Similarity 66.7%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 YOL 4  
 DB 1 YPL 3

## RESULT 11

J01273  
 C:Species: Anthopleura elegantissima  
 C:Date: 31-Mar-1992 #sequence\_revision 04-Dec-1992 #text\_change 08-Dec-1995  
 C:Accession: J01273  
 R:Notackker, H.P.; Rinehart, K.L.; Grimmelshuis, C.J.P.  
 Biochem. Biophys. Res. Commun. 179, 1205-1211, 1991  
 A:Title: Isolation of L-3-phenylalanyl-phe-lys-Ala-NH2 (Antho-Kamide), a novel neuro  
 A:Reference number: J01273; MUID:92028852  
 A:Accession: J01273  
 A:Molecule type: protein  
 A:Residues: 1-4 <NOT>  
 C:Comment: The carboxyl-terminal amide probably arises from cleavage of a following g  
 C:Keywords: amidated carboxyl end; neuropeptide; phenylacetylation  
 F:1/Modified site: L-3-phenylalanyl acid (Phe) #status experimental  
 F:4/Modified site: amidated carboxyl end (Ala) #status experimental

Query Match 28.6%; Score 10; DB 2; Length 4;  
 Best Local Similarity 33.3%; Pred. No. 2.8e+05;  
 Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FYO 3  
 DB 1 FPK 3

RESULT 12  
 S60293  
 C:Species: Drosophila melanogaster  
 C:Date: 19-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 17-Mar-1999  
 C:Accession: S60293  
 R:Chapel, S.; Sobrier, M.L.; Montplied, P.; Micard, D.; Brunat, A.; Couderc, J.L.; Das  
 Insect Mol. Biol. 2, 39-48, 1993  
 A:Title: In Drosophila Kc cells 20-OHE induction of the 60c beta-3 tubulin gene expre  
 A:Reference number: S60292; MUID:97242543  
 A:Accession: S60293  
 A:Molecule type: DNA  
 A:Residues: 1-6 <CHM>  
 A:Cross-references: EMBL:X60394  
 C:Genetics:  
 A:Gene: FlyBase:fbgr  
 A:Cross-references: FlyBase:FBgn0003888  
 A:introns: 4/1

Query Match 28.6%; Score 10; DB 2; Length 6;  
 Best Local Similarity 66.7%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 AIT 7  
 DB 4 AVT 6

RESULT 13  
 B35640  
 C:Species: Mus musculus (house mouse)  
 C:Date: 28-Sep-1990 #sequence\_revision 28-Sep-1990 #text\_change 24-Jun-1993  
 C:Accession: B35640  
 R:Chen, Y.T.; Rettig, W.J.; Yenamandra, A.K.; Kozak, C.A.; Chaganti, R.S.K.; Posner,  
 Proc. Natl. Acad. Sci. U.S.A. 87, 3077-3081, 1990  
 A:Title: Cerebellar degeneration-related antigen: a highly conserved neuroectodermal  
 A:Reference number: A35640; MUID:90222173  
 A:Accession: B35640  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-6 <CHE>

Query Match 28.6% Score 10; DB 2; Length 6;  
 Best Local Similarity 33.3%; Pred. No. 2.8e+05;  
 Matches 1; Conservative 2; Mismatches 0; Indels 0;

OY 1 FYQ 3  
 1:  
 Db 1 FME 3

## RESULT 14

I49421  
 Iaminin B1 - western wild mouse (fragment)  
 C:Species: Mus spretus (western wild mouse)  
 C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
 C:Accession: I49421  
 R:KO, M.S.; Wang, X.; Horton, J.H.; Hagen, M.D.; Takahashi, N.; Maezaki, Y.; Nadeau, J.H.  
 Mamm. Genome 5, 349-355, 1994  
 A>Title: Genetic mapping of 40 cDNA clones on the mouse genome by PCR.  
 A:Reference number: 148934; MUID:94319082  
 A:Accession: I49421  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-6 <RES>  
 A:Cross-references: EMBL:U05736; NID:g497073; PIDN:AA860477.1; PID:g642829

Query Match 28.6% Score 10; DB 2; Length 6;  
 Best Local Similarity 40.0%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 0; Mismatches 3; Indels 0;

OY 2 YQAL 6  
 1:  
 Db 2 YSTCL 6

## RESULT 15

A43129  
 neuropeptide GnFRFamide - tapeworm (Montezia expansa)  
 C:Species: Montezia expansa  
 C>Date: 10-Nov-1997 #sequence\_revision 14-Nov-1997 #text\_change 14-Nov-1997  
 C:Accession: A43129  
 R:Maule, A.; Shaw, C.; Halton, D.; Thim, L.  
 Biochem. Biophys. Res. Commun. 193, 1054-1060, 1993  
 A>Title: GnFRFamide: A novel FMRFamide-immunoreactive peptide isolated from the sheep  
 A:Reference number: A43129; MUID:93312289  
 A:Accession: A43129  
 A:Molecule type: protein  
 A:Residues: 1-6 <MAU>  
 C:Keywords: amidated carboxyl end; neuropeptide  
 F:6/Modified site: amidated carboxyl end (Phe) #status predicted

Query Match 28.6% Score 10; DB 2; Length 6;  
 Best Local Similarity 33.3%; Pred. No. 2.8e+05;  
 Matches 1; Conservative 2; Mismatches 0; Indels 0;

OY 1 FYQ 3  
 1:  
 Db 3 FFR 5

Search completed: July 15, 2002, 13:27:08  
 Job time: 438 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:38:10 ; Search time 24.49 Seconds

(without alignments)  
11,067 Million cell updates/sec

Title: US-09-712-819A-3

Perfect score: 35

Sequence: 1 FYOLALT 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 84

Minimum DB seq length: 0

Maximum DB seq length: 7

Post-processing: Minimum Match 08

Maximum Match 1008

Database : SwissProt.40.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	45.7	4	1	FYRI_ANCEL
2	12	34.3	5	1	UC22_MAIZE
3	11	31.4	5	1	AL14_CARMA
4	11	31.4	7	1	AL17_CYPDO
5	11	31.4	7	1	GFPR_MOUSE
6	10	28.6	4	1	FFFR_ANCEL
7	10	28.6	6	1	FFFR_MONEX
8	10	28.6	7	1	AL12_CARMA
9	10	28.6	7	1	AL13_CARMA
10	10	28.6	7	1	AL15_CARMA
11	10	28.6	7	1	AL16_CARMA
12	9	25.7	5	1	PAR2_PARMA
13	9	25.7	5	1	RE11_LITRU
14	9	25.7	5	1	RE21_LITRU
15	9	25.7	5	1	RE31_LITRU
16	9	25.7	5	1	RE32_LITRU
17	9	25.7	7	1	E105_LITRU
18	9	25.7	7	1	TY51_LITRU
19	9	25.7	7	1	WM42_ACHFU
20	8	23.9	6	1	OVW2_LEPDE
21	8	23.9	7	1	CIA_ENTFA
22	8	22.9	7	1	HT7_PIG
23	8	22.9	7	1	MNP1_LEPDE
24	7	20.0	4	1	FAR3_HIRME
25	7	20.0	4	1	FAR4_HIRME
26	7	20.0	4	1	RM01_YEAST
27	7	20.0	5	1	BIOA_CITFR
28	7	20.0	5	1	E104_LITRU
29	7	20.0	5	1	FARF_AATTR
30	7	20.0	5	1	PRCT_PERAM
31	7	20.0	5	1	PSK_DAVCA
32	7	20.0	6	1	ASP2_LACSN
33	7	20.0	7	1	CARP_MYTED

## ALIGNMENTS

RESULT	ID	FYRI_ANCEL	STANDARD	PRT	4 AA
AC	P58706				
DT	01-MAR-2002 (Rel. 41, last sequence update)				
DT	01-MAR-2002 (Rel. 41, last annotation update)				
DE	Antho-Riamide I [contains: Antho-Riamide III]				
OS	Anthopleura elegantissima (See anemone)				
OC	Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinoptaria;				
OC	Nysanthese; Metazoa; Cnidaria; Anthozoa; Anthopleura.				
OX	NCBI_TaxID=6110;				
RN	[1]				
RP	SEQUENCE.				
RX	PubMed=1821096;				
RA	Nothacker H.-P., Rinehart K.L. Jr., McFarlane I.D.,				
RA	Grimmelikhuizen C.J.P.;				
RT	*Isolation of two novel neuropeptides from sea anemones; the unusual,				
RT	biologically active L-3-phenylacetyl-Tyr-Arg-Ile-NH2 and its				
RT	Peptides 12:1165-1173(1991).				
RN	[2]				
RP	FUNCTION.				
RX	PubMed=8397415;				
RA	McFarlane I.D., Hudman D., Nothacker H.-P., Grimmelikhuizen C.J.P.;				
RT	*The expansion behaviour of sea anemones may be coordinated by two				
RT	inhibitory neuropeptides, Antho-Kamide and Antho-Riamide.*;				
RT	Proc. R. Soc. Lond., B, Biol. Sci. 253:183-188(1993).				
CC	-/- FUNCTION: Inhibits spontaneous contractions in several muscle				
CC	groups. May be involved in the expansion phase of feeding				
CC	behaviour in sea anemones.				
CC	-/- SUBCELLULAR LOCATION: Secreted.				
CC	-/- TISSUE SPECIFICITY: Neuron-specific.				
KW	Neuropeptide; Amidation.				
FT	CHAIN	1	4		ANTHO-RIAMIDE I.
FT	CHAIN	2	4		ANTHO-RIAMIDE II.
FT	MOD_RES	1	1		L-3-PHENYLACETYL.
FT	MOD_RES	4	4		AMIDATION.
SEQ	SEQUENCE	4 AA: 598 MW: 60441B59A000000 CRC64;			

Query Match 45.7%; Score 16; DB 1; Length 4;  
Best Local Similarity 50.0%; Pred. No. 16+05;  
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY	1	FYOL	4	
DB	1	FYRI	4	

RESULT 2  
UC22\_MAIZE STANDARD; PRT; 5 AA.  
ID UC22\_MAIZE  
AC P80628;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUL-1999 (rel. 38, Last annotation update)  
 DE Unknown protein from 2D-page of eliolated coleoptile (Spot 474)  
 DE (Fragment).  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Panicoideae; Andropogoneae; Zea.  
 OX NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=coleoptile;  
 RA Toulzet P., Riccardi F., Morin C., Dameray C., Huet J.-C.,  
 RA Fernollet J.-C., Zivy M., de Vienne D.;  
 RT "The maize two dimensional gel protein database: towards an integrated  
 RT genome analysis program."  
 RL Theor. Appl. Genet. 93:997-1005(1996).  
 CC -1- MISCELLANEOUS: ON THE 2D-GEL, THE DETERMINED PT OF THIS UNKNOWN  
 CC PROTEIN IS: 6.1, ITS MW IS: 30.4 kDa.  
 DR MAIZE-2DPAGE; P80628; COLEOPTILE.  
 DR MAIZE2DB; 123954;  
 DR NON\_TER 1 1  
 FT NON\_TER 5 5  
 SQ SEQUENCE 5 AA: 654 MW: 72CB19C9C0300000 CRC64:

Query Match 34.38; Score 12; DB 1; Length 5;  
 Best Local Similarity 25.08; Pred. No. 1e+05; Indels 0; Gaps 0;  
 Matches 1; Conservative 3; Mismatches 0;

OY 1 YQL 4  
 DB 2 FFV 5

RESULT 3  
 ALL7\_CARMA STANDARD; PRT: 5 AA.  
 ID AL14\_CARMA  
 AC P81817;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Carcinus maenas (Common shore crab) (Green crab).  
 OS Carcinus maenas (Common shore crab) (Green crab).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
 OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
 OX NCBI\_TaxID=6759;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
 RX MEDLINE=98121193; PubMed=9461265;  
 RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jiros P.P.,  
 RA Thorpe A.;  
 RT "Isolation and identification of multiple neuropeptides of the  
 RT allatostatin superfamily in the shore crab Carcinus maenas."  
 RL Eur. J. Biochem. 250:727-734(1997).  
 CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 CC Neuropeptide; Amidation; Multigene family.  
 FT MOD\_RES 5 5  
 FT INT\_MET 5 5  
 SQ SEQUENCE 5 AA: 586 MW: 672879D5AB300000 CRC64:

Query Match 31.48; Score 11; DB 1; Length 5;  
 Best Local Similarity 40.08; Pred. No. 1e+05; Indels 0; Gaps 0;  
 Matches 2; Conservative 0; Mismatches 3;

OY 2 YQL 6  
 DB 1 YSFG 5

RESULT 4

ALL7\_CYPDO STANDARD; PRT: 7 AA.  
 ID ALL7\_CYPDO  
 AC P82158;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Cydia pomonella (Coddling moth).  
 OS Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Tortricidae; Tortricidae; Olethreutinae; Cydia.  
 OX NCBI\_TaxID=82600;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Larva;  
 RA MEDLINE=98054539; PubMed=9392829;  
 RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,  
 RA Davey M., East P.D., Thorpe A.;  
 RT "Lepidopteran peptides of the allatostatin superfamily."  
 RL Peptides 18:1301-1309(1997).  
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 CC Neuropeptide; Amidation.  
 FT MOD\_RES 7 7  
 SQ SEQUENCE 7 AA: 873 MW: 672879CABB569350 CRC64:

Query Match 31.48; Score 11; DB 1; Length 7;  
 Best Local Similarity 40.08; Pred. No. 1e+05; Indels 0; Gaps 0;  
 Matches 2; Conservative 0; Mismatches 3;

OY 2 YQL 6  
 DB 3 YDFGL 7

RESULT 5  
 GPRP\_MOUSE STANDARD; PRT: 7 AA.  
 ID GPRP\_MOUSE  
 AC P99025;  
 DT 13-DEC-1998 (Rel. 37, Created)  
 DT 13-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE GTP cyclohydrolase I feedback regulatory protein (P35) (Fragment).  
 GN GCHFR OR GFRP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Liver;  
 RA Sanchez J.-C., Rouge V., Frutiger S., Hughes G., Yan J.X.,  
 RA Hoogland C., Appel R.D., Binz P.-A., Hochstrasser D.F.,  
 RA Cowthorne M.;  
 RT "Function (AUG-1998) to the SWISS-PROT data bank.  
 RT allatostatin superfamily in the shore crab Carcinus maenas."  
 RL Eur. J. Biochem. 250:727-734(1997).  
 CC -1- FUNCTION: MEDIATES TETRAHYDROBIOPPERIN INHIBITION OF GTP  
 CC CYCLOHYDROLASE I. THIS INHIBITION IS REVERSED BY L-PHENYLANILINE  
 CC (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC SWISS-2DPAGE; P99025; MOUSE.  
 DR INT\_MET 7 7  
 FT INT\_MET 7 7  
 SQ SEQUENCE 7 AA: 806 MW: 71B5B057273B4700 CRC64:

Query Match 31.48; Score 11; DB 1; Length 7;  
 Best Local Similarity 50.08; Pred. No. 1e+05; Indels 0; Gaps 0;  
 Matches 3; Conservative 0; Mismatches 3;

OY 2 YQL 7  
 DB 2 YLIST 7

```

RESULT 6
FFKA_AMEL STANDARD: PRT: 4 AA.
ID FFKA_AMEL
AC P58705: 2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Antho-Kamide.
OS Anthopleura elegantissima (Sea anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinaria;
OC Nynanthaeae; Actiniidae; Anthopleura.
OX NCBI_TaxID=6110;
RN [1]
RP SEQUENCE.
RX PubMed:1681803; Rinehart K.L. Jr., Grimmelikhuijzen C.J.P.;
RA Notackner H.-P., Rinehart K.L. Jr., Grimmelikhuijzen C.J.P.;
RT "Isolation of L-3-phenyllactyl-Phe-Lys-Ala-NH2 (Antho-Kamide), a
RT novel neuropeptide from sea anemones."
RL Biochem. Biophys. Res. Commun. 179:1205-1211(1991).
RN [2]
RP FUNCTION.
RX PubMed:8397415;
RA McFarlane I.D., Hudman D., Notackner H.-P., Grimmelikhuijzen C.J.P.;
RT "The expansion behaviour of sea anemones may be coordinated by two
RT inhibitory neuropeptides, Antho-Kamide and Antho-Ramide."
RL Proc. R. Soc. Lond. B, Biol. Sci. 253:183-188(1993).
CC -1- FUNCTION: Inhibits spontaneous contractions in several muscle
CC groups. May be involved in the expansion phase of feeding
CC behaviour in sea anemones.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Neuron-specific.
KW Neuropeptide; Amidation.
FT MOD_RES 1 1 L-3-PHENYLLACTYL.
FT MOD_RES 4 4 AMIDATION.
SQ SEQUENCE 4 AA: 512 MW: 6D339C9A0D000000 CRC64;

Query Match 28.6%; Score 10; DB 1; Length 4;
Best Local Similarity 33.3%; Pred. No. 1e+05;
Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FYQ 3
DB 1 FPK 3

RESULT 7
FARP_MONEX STANDARD: PRT: 6 AA.
ID FARP_MONEX
AC P41966:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FMRamide-like neuropeptide GNFRR-amide.
OS Monieza expansa (Sheep tapeworm).
OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
OC Rhadithophora; Eutelethophora; Revertosporina; Medusozoa;
OC Neodermata; Cestoda; Eucestoda; Cyclophyllidae; Aplocephalidae;
OX NCBI_TaxID=28841;
RN [1]
RP SEQUENCE.
RX MEDLINE:93312289; PubMed:833531;
RA Maule A.G., Shaw C., Hulton D.W., Thim L.;
RT "GNFRRamide: a novel FMRamide-immunoreactive peptide isolated from
RT the sheep tapeworm, Monieza expansa."
RL Biochem. Biophys. Res. Commun. 193:1054-1060(1992).
CC -1- SIMILARITY: BELONGS TO THE FARP (FMRAMIDE RELATED PEPTIDE)
CC FAMILY.
KW Neuropeptide; Amidation.
FT MOD_RES 6 6 AMIDATION.
SQ SEQUENCE 6 AA: 787 MW: 69DA09C9C4481000 CRC64;

```

```

Query Match 28.6%; Score 10; DB 1; Length 6;
Best Local Similarity 33.3%; Pred. No. 1e+05;
Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FYQ 3
DB 3 FPR 5

RESULT 8
ALL2_CARMA STANDARD: PRT: 7 AA.
ID ALL2_CARMA
AC P81805:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Carcinustatin 2.
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RX MEDLINE:98121193; PubMed:9461295;
RA Duvre H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA Thorpe A.;
RT "Isolation and identification of multiple neuropeptides of the
RT allatostatin superfamily in the shore crab Carcinus maenas."
RL Eur. J. Biochem. 250:727-734(1997).
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
KW Neuropeptide; Amidation; Multigene family.
FT MOD_RES 7 7 AMIDATION (POTENTIAL).
SQ SEQUENCE 7 AA: 770 MW: 672879CDB5DBD70 CRC64;

Query Match 28.6%; Score 10; DB 1; Length 7;
Best Local Similarity 40.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 YQIAL 6
DB 3 YAFGL 7

RESULT 9
ALL3_CARMA STANDARD: PRT: 7 AA.
ID ALL3_CARMA
AC P81806:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Carcinustatin 3.
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RX MEDLINE:98121193; PubMed:9461295;
RA Duvre H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA Thorpe A.;
RT "Isolation and identification of multiple neuropeptides of the
RT allatostatin superfamily in the shore crab Carcinus maenas."
RL Eur. J. Biochem. 250:727-734(1997).
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.

```

KW Neuropeptide; Multigene family.  
SQ SEQUENCE 7 AA: 796 MW: 672879CDCB476B70 CRC64:

Query Match 28.6%; Score 10; DB 1; Length 7;  
Best Local Similarity 40.0%; Pred. No. 1e+05;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 YOTAL 6  
DB 3 YAFGL 7

RESULT 10  
ALL4\_CARMA STANDARD; PRT; 7 AA.  
ID ALL4\_CARMA  
AC P81807;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Carcinus statin 4.  
OS Carcinus maenas (Common shore crab) (Green crab).  
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
OX NCBI\_TaxID=6759;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
RX MEDLINE=98121193; PubMed=9461295;  
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
Thorpé A.;  
RT Isolation and identification of multiple neuropeptides of the  
RT allatostatins superfamily in the shore crab Carcinus maenas.";  
RL Eur. J. Biochem. 250:727-734(1997).  
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
KW Neuropeptide; Multigene family.  
SQ SEQUENCE 7 AA: 782 MW: 672879CDCB476AC0 CRC64:

Query Match 28.6%; Score 10; DB 1; Length 7;  
Best Local Similarity 40.0%; Pred. No. 1e+05;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 YOTAL 6  
DB 3 YAFGL 7

RESULT 11  
ALL5\_CARMA STANDARD; PRT; 7 AA.  
ID ALL5\_CARMA  
AC P81808;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Carcinus statin 5.  
OS Carcinus maenas (Common shore crab) (Green crab).  
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
OX NCBI\_TaxID=6759;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
RX MEDLINE=98121193; PubMed=9461295;  
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
Thorpé A.;  
RT Isolation and identification of multiple neuropeptides of the  
RT allatostatins superfamily in the shore crab Carcinus maenas.";  
RL Eur. J. Biochem. 250:727-734(1997).  
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.

CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
KW Neuropeptide; Amidation; Multigene family.  
FT MOD\_RES 7  
SQ SEQUENCE 7 AA: 781 MW: 672879CDCB476420 CRC64:

Query Match 28.6%; Score 10; DB 1; Length 7;  
Best Local Similarity 40.0%; Pred. No. 1e+05;  
Matches 2; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 YOTAL 6  
DB 3 YAFGL 7

RESULT 12  
PAP2\_PARMA STANDARD; PRT; 5 AA.  
ID PAP2\_PARMA  
AC P81864;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Pardaxin II (PxiI) (Fragment).  
OS Pardachirus marmoratus (Red sea moose sole).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;  
OC Soleoidae; Soleidae; Pardachirus.  
OX NCBI\_TaxID=31087;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=SKIN secretion;  
RX MEDLINE=87057369; PubMed=3782138;  
RA Lazarovici P., Primor N., Loew L.M.;  
RT "Purification and pore-forming activity of two hydrophobic  
RT polypeptides from the secretion of the Red sea moose sole (Pardachirus  
RT marmoratus).";  
RL J. Biol. Chem. 261:16704-16713(1986).  
CC -1- FUNCTION: EXHIBITS UNUSUAL SHARK REPLENT AND SURFACTANT  
CC PROPERTIES. FORMS VOLTAGE-DEPENDENT, ION-PERMEABLE CHANNELS  
CC IN MEMBRANES. AT HIGH CONCENTRATION CAUSES CELL MEMBRANE LYSIS.  
CC -1- SUBUNIT: MONOMER. IN AQUEOUS SOLUTION EXISTS AS A Tetramer.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: BELONGS TO THE PARDAXIN FAMILY.  
KW Toxin.  
FT NON TER  
SQ SEQUENCE 5 AA: 614 MW: 7769C9C9C8100000 CRC64:

Query Match 25.7%; Score 9; DB 1; Length 5;  
Best Local Similarity 50.0%; Pred. No. 1e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FY 2  
DB 2 FF 3

RESULT 13  
REL1\_LITRU STANDARD; PRT; 5 AA.  
ID REL1\_LITRU  
AC P82070;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Rubellidin 1.1.  
OS Iliorbia rubella (Desert tree frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Urotrita.  
OX NCBI\_TaxID=104895;  
RN [1]  
RP SEQUENCE, AND MASS SPECTROMETRY.

RC TISSUE-Skin secretion;  
 RA Steinboerner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,  
 Tyler M.J., Wallace J.C.;  
 RT "The structure of new peptides from the Australian red tree frog  
 'Litoria rubella', the skin peptide profile as a probe for the study  
 of evolutionary trends of amphibians.";  
 RL Aust. J. Chem. 49:955-963(1996).  
 CC -1- FUNCTION: SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC  
 CC ACTIVITY.  
 CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
 CC -1- MASS SPECTROMETRY: MW=598; METHOD-FAB.  
 KW Amphibian skin.  
 SQ SEQUENCE 5 AA: 598 MW: 60D9C9CBA2A00000 CRC64;

Query Match 25.7%; Score 9; DB 1; Length 5;  
 Best Local Similarity 50.0%; Pred. No. 1e+05;  
 Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FY 2  
 DB 3 FF 4

RESULT 14  
 ID RE21\_LITRU STANDARD: PRT: 5 AA.  
 AC P82071;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Rubellidin 2.1.  
 OS Litoria rubella (Desert tree frog).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
 OC Litoria.  
 OX NCBI\_TaxID=104895;  
 RN [1]  
 RP SEQUENCE AND MASS SPECTROMETRY.  
 RC TISSUE-Skin secretion;  
 RA Steinboerner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,  
 Tyler M.J., Wallace J.C.;  
 RT "The structure of new peptides from the Australian red tree frog  
 'Litoria rubella', the skin peptide profile as a probe for the study  
 of evolutionary trends of amphibians.";  
 RL Aust. J. Chem. 49:955-963(1996).  
 CC -1- FUNCTION: SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC  
 CC ACTIVITY.  
 CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
 CC -1- MASS SPECTROMETRY: MW=626; METHOD-FAB.  
 KW Amphibian skin.  
 SQ SEQUENCE 5 AA: 626 MW: 6D9C9CBA10300000 CRC64;

Query Match 25.7%; Score 9; DB 1; Length 5;  
 Best Local Similarity 50.0%; Pred. No. 1e+05;  
 Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FY 2  
 DB 3 FF 4

RESULT 15  
 ID RE31\_LITRU STANDARD: PRT: 5 AA.  
 AC P82072;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Rubellidin 3.1.  
 OS Litoria rubella (Desert tree frog).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
 OC Litoria.  
 OX NCBI\_TaxID=104895;  
 RN [1]  
 RP SEQUENCE AND MASS SPECTROMETRY.  
 RC TISSUE-Skin secretion;  
 RA Steinboerner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,  
 Tyler M.J., Wallace J.C.;  
 RT "The structure of new peptides from the Australian red tree frog  
 'Litoria rubella', the skin peptide profile as a probe for the study  
 of evolutionary trends of amphibians.";  
 RL Aust. J. Chem. 49:955-963(1996).  
 CC -1- FUNCTION: SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC  
 CC ACTIVITY.  
 CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
 CC -1- MASS SPECTROMETRY: MW=655; METHOD-FAB.  
 KW Amphibian skin; Amadation.  
 FT MOD\_BRS 5  
 SQ SEQUENCE 5 AA: 656 MW: 71A9C9CBA10300000 CRC64;

Query Match 25.7%; Score 9; DB 1; Length 5;  
 Best Local Similarity 50.0%; Pred. No. 1e+05;  
 Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FY 2  
 DB 3 FF 4

Search completed: July 15, 2002, 13:38:10  
 Job time: 710 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:37:39 ; Search time 86.73 Seconds

(without alignments)  
13.962 Million cell updates/sec

Title: US-09-712-819A-3  
Sequence: 35  
1 FYQALIR 7

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994923 residues

Total number of hits satisfying chosen parameters: 65

Minimum DB seq length: 0  
Maximum DB seq length: 7

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SPREMBL\_19:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_unclassified:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeop:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12	34.3	7	8 099182	099182 gnatholebia
2	10	28.6	7	2 P72081	P72081 nocardia la
3	9	25.7	5	13 P82070	P82070 litorea rub
4	9	25.7	5	13 P82071	P82071 litorea rub
5	9	25.7	5	13 P82072	P82072 litorea rub
6	9	25.7	5	13 P82073	P82073 litorea rub
7	9	25.7	7	12 09Y010	09Y010 transmissib
8	9	25.7	7	13 P82065	P82065 litorea rub
9	9	25.7	7	13 P82101	P82101 litorea rub
10	8	22.9	7	8 095945	095945 saccharomyc
11	8	22.9	7	10 09C283	09C283 arabidopsids
12	8	22.9	7	11 063668	063668 ratius norv
13	8	22.9	7	12 066113	066113 cherry leaf
14	7	20.0	5	13 P82100	P82100 litorea rub
15	7	20.0	6	10 P82181	P82181 spinacia ol
16	7	20.0	6	10 P82182	P82182 spinacia ol

17	7	20.0	7	2 007354	007354 synecococc
18	7	20.0	7	2 047029	047029 enterobacte
19	7	20.0	7	2 050556	050556 actinobacti
20	7	20.0	7	2 034028	034028 spingomonas
21	7	20.0	7	2 054248	054248 streptomyc
22	7	20.0	7	4 015897	015897 homo sapien
23	7	20.0	7	10 P93233	P93233 lycopersico
24	7	20.0	7	12 09Y180	09Y180 human adeno
25	7	20.0	7	12 09Y109	09Y109 human adeno
26	7	20.0	7	12 09Y193	09Y193 human adeno
27	6	17.1	5	10 09Y007	09Y007 hordenum vul
28	6	17.1	5	13 P82099	P82099 litorea rub
29	6	17.1	6	13 P82096	P82096 litorea rub
30	6	17.1	7	4 015903	015903 homo sapien
31	6	17.1	7	8 098866	098866 spinacia ol
32	6	17.1	7	10 P82445	P82445 nicotiana t
33	6	17.1	7	13 042564	042564 figu rubrip
34	6	17.1	7	15 007624	007624 rous sarcom
35	5	14.3	5	2 P83073	P83073 bacillus ce
36	5	14.3	5	10 P82541	P82541 spinacia ol
37	5	14.3	7	2 047477	047477 escherichia
38	5	14.3	7	2 047505	047505 escherichia
39	5	14.3	7	2 P70804	P70804 azotobacter
40	5	14.3	7	6 028742	028742 oryctolagus
41	5	14.3	7	8 P92421	P92421 psathyrosta
42	5	14.3	7	8 P92385	P92385 hordenum mar
43	5	14.3	7	8 P92372	P92372 haynaldia v
44	5	14.3	7	8 P92403	P92403 lophopyrgm
45	5	14.3	7	8 P92425	P92425 pseudorocgn

## ALIGNMENTS

RESULT 1  
ID 099182 PRELIMINARY: PRT: 7 AA.  
AC 099182;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE CYTOCHROME OXIDASE I (FRAGMENT).  
GN COI  
OS Gnatholebias zonatus.  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorphae; Acanthopterygii; Percomorphae; Atherinomorpha;  
OC Cyprinodontiformes; Aplocheilidae; Rivulinae; Gnatholebias.  
OX NCBI\_TaxID=135316;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20072928; PubMed=10603257;  
RA Murphy W.J., Thomerson J.E., Collier G.E.;  
RT "Phylogeny of the Neotropical Killifish Family Rivulidae  
RT (Cyprinodontiformes, Aplocheilidae) inferred from mitochondrial DNA  
RT sequences".  
RL Mol. Phylogenet. Evol. 13:289-301(1999).  
DR EMBL; AF002591; AAD01074.1; -.  
FT Mitochondrion.  
FT NON\_TBR  
SQ SEQUENCE 7 AA; 899 MW; 672721FCB572030 CRC64;

Query Match 34.3%; Score 12; DB 8; Length 7;  
Best Local Similarity 100.0%; Pred No. 5.6e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 YQ 3  
II  
Db 3 YQ 4

```

RESULT 2
P72081 PRELIMINARY; PRT; 7 AA.
AC P72081;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 3'-METHYLCEPHEM HYDROXYLASE (FRAGMENT).
GN CEF3.
OS Nocardia lactamdurans.
OC Actinomycetales; Pseudonocardineae; Pseudonocardiales; Amycolatopsis.
OC Actinomycetales; Pseudonocardineae; Pseudonocardiales; Amycolatopsis.
OX NCBI_TaxID=1913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96009872; PubMed=7557411;
RA Coque J., Perez-Llarena F.J., Enguita J.F., Fuente J.L., Martin J.F.,
RA Liras P.;
RT "Characterization of the cmcH genes of Nocardia lactamdurans and
RT Streptomyces clavuligerus encoding a functional 3'-hydroxymethylcephem
RT O-carbamoyltransferase for cephamycin biosynthesis."
RL Gene 162:21-27(1995).
DR EMBL: Z21682; CAA79797.1; -.
FT NON_TER 1
SQ SEQUENCE 7 AA; 746 MW; 75A1B2CDD1E681C0 CRC64;

Query Match 28.6%; Score 9; DB 13; Length 7;
Best Local Similarity 66.7%; Pred. No. 5.6e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 5 ALT 7
DB 4 AVT 6

RESULT 3
P82070 PRELIMINARY; PRT; 5 AA.
AC P82070;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE RUBELLIDIN 1.1.
OS Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE-SKIN SECRETION;
RA Steinboerner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,
RA Tyler M.J., Wallace J.C.;
RT "The structure of new peptides from the Australian red tree frog
RT 'Litoria rubella', the skin peptide profile as a probe for the study
RT of evolutionary trends of amphibians."
RL Aust. J. Chem. 49:955-963(1996).
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR
CC ANTI-BIOTIC ACTIVITY.
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.
CC -1- MASS SPECTROMETRY: MM=538; METHOD=FAB.
KW Amphibian skin.
SQ SEQUENCE 5 AA; 598 MW; 6DD9C9CAB2A00000 CRC64;

Query Match 25.7%; Score 9; DB 13; Length 5;
Best Local Similarity 50.0%; Pred. No. 5.6e+05;
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FY 2
DB 3 FF 4

```

```

RESULT 4
P82071 PRELIMINARY; PRT; 5 AA.
AC P82071;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE RUBELLIDIN 2.1.
OS Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE-SKIN SECRETION;
RA Steinboerner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,
RA Tyler M.J., Wallace J.C.;
RT "The structure of new peptides from the Australian red tree frog
RT 'Litoria rubella', the skin peptide profile as a probe for the study
RT of evolutionary trends of amphibians."
RL Aust. J. Chem. 49:955-963(1996).
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR
CC ANTI-BIOTIC ACTIVITY.
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.
CC -1- MASS SPECTROMETRY: MM=626; METHOD=FAB.
KW Amphibian skin.
SQ SEQUENCE 5 AA; 626 MW; 6DD9C9CB10300000 CRC64;

Query Match 25.7%; Score 9; DB 13; Length 5;
Best Local Similarity 50.0%; Pred. No. 5.6e+05;
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FY 2
DB 3 FF 4

RESULT 5
P82072 PRELIMINARY; PRT; 5 AA.
AC P82072;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE RUBELLIDIN 3.1.
OS Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE-SKIN SECRETION;
RA Steinboerner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,
RA Tyler M.J., Wallace J.C.;
RT "The structure of new peptides from the Australian red tree frog
RT 'Litoria rubella', the skin peptide profile as a probe for the study
RT of evolutionary trends of amphibians."
RL Aust. J. Chem. 49:955-963(1996).
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR
CC ANTI-BIOTIC ACTIVITY.
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.
CC -1- MASS SPECTROMETRY: MM=655; METHOD=FAB.
KW Amphibian skin; Amidation.
FT MOD_RES 5
SQ SEQUENCE 5 AA; 656 MW; 71A9C9CB10300000 CRC64;

Query Match 25.7%; Score 9; DB 13; Length 5;

```

Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FY 2  
DB 3 FF 4

RESULT 6  
P82073 PRELIMINARY; PRT; 5 AA.  
AC P82073:  
DT 01-MAY-2000 (TREMUREL. 13, Created)  
DT 01-MAY-2000 (TREMUREL. 13, Last sequence update)  
DT 01-MAY-2000 (TREMUREL. 13, Last annotation update)  
DE RUBELLIDIN 3.2.  
OS Litorea rubella (Desert tree frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litorea.  
OX NCBI\_TaxID=104895;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=SKIN SECRETION;  
RA Weinreb P.A., Bowie J.H., Tyler M.J., Wallace J.C.;  
RT "Peptides from the skin glands of the Australian buzzing tree frog  
Litorea rubella, comparison with the skin peptides from Litorea  
rubella.";  
RU Aust. J. Chem. 52:0-0(1999).  
CC -1- FUNCTION: CERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR  
ANTIBIOTIC ACTIVITY.  
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
KW Amphibian skin.  
SQ SEQUENCE 5 AA; 570 MW; 71A9C9C862A0000 CRC64;

Query Match 25.7%; Score 9; DB 13; Length 5;  
Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FY 2  
DB 3 FF 4

RESULT 7  
O9Y010 PRELIMINARY; PRT; 7 AA.  
ID O9Y010:  
AC O9Y010:  
DT 01-MAY-1999 (TREMUREL. 10, Created)  
DT 01-MAY-1999 (TREMUREL. 10, Last sequence update)  
DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)  
DE HYPOTHECAL FUSION PROTEIN.  
OC Transmissible gastroenteritis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxID=11449;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90909045; PubMed=9882359;  
RA Ireta A., Smerdou C., Alonso S., Penzes Z., Mendez A., Plana-Duran J.,  
RJ Enjuanes L.;  
RT "Replication and packaging of transmissible gastroenteritis  
coronavirus-derived synthetic minigenomes.";  
RL J. Virol. 73:1535-1545(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95159435; PubMed=7856095;  
RA Rieunet J., Rasschaert D., Lambert P., Levy L., Vende P., Laude H.;  
RT "Complete sequence (20 kilobases) of the polypeptide-encoding gene 1  
of transmissible gastroenteritis virus.";  
RL Virology 206:817-822(1995).  
RN [3]

RP SEQUENCE FROM N.A.  
RX MEDLINE=88078100; PubMed=2825819;  
RA Rasschaert D., Gellif J., Laude H.;  
RT "Enteric coronavirus TGEV: partial sequence of the genomic RNA lts  
organization and expression.";  
RL Biochimie 69:591-600(1987).  
DR EMBL: AJ011482; CA009625.1;  
SQ SEQUENCE 7 AA; 927 MW; 69D6D7273B5726F0 CRC64;

Query Match 25.7%; Score 9; DB 12; Length 7;  
Best Local Similarity 66.7%; Pred. No. 5.6e+05;  
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FY 4  
DB 3 YL 5

RESULT 8  
P82065 PRELIMINARY; PRT; 7 AA.  
ID P82065:  
AC P82065:  
DT 01-MAY-2000 (TREMUREL. 13, Created)  
DT 01-MAY-2000 (TREMUREL. 13, Last sequence update)  
DT 01-MAY-2000 (TREMUREL. 13, Last annotation update)  
DE LITOPHYLLIN 5.1.  
OS Litorea rubella (Desert tree frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litorea.  
OX NCBI\_TaxID=104895;  
RN [1]  
RP SEQUENCE, AND MASS SPECTROMETRY.  
RC TISSUE=SKIN SECRETION;  
RA Steinborner S.T., Weinreb P.A., Waugh R.J., Bowie J.H., Gao C.,  
RA Tyler M.J., Wallace J.C.;  
RT "The structure of new peptides from the Australian red tree frog  
Litorea rubella, the skin peptide profile as a probe for the study  
of evolutionary trends of amphibians.";  
RU Aust. J. Chem. 49:955-963(1996).  
CC -1- FUNCTION: MAY ACT AS A NEUROMODULATOR OR NEUROTRANSMITTER.  
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
CC -1- MASS SPECTROMETRY: MW=965; METHOD=HAB.  
KW Amphibian skin; Amidation; Neuropeptide.  
FT MOD\_RES 1 1 PYROLIDONE CARBOXYLIC ACID.  
FT MOD\_RES 7 7 AMIDATION.  
SQ SEQUENCE 7 AA; 983 MW; 7401E9D3676046B0 CRC64;

Query Match 25.7%; Score 9; DB 13; Length 7;  
Best Local Similarity 33.3%; Pred. No. 5.6e+05;  
Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 FY 3  
DB 5 FHR 7

RESULT 9  
P82101 PRELIMINARY; PRT; 7 AA.  
ID P82101:  
AC P82101:  
DT 01-MAY-2000 (TREMUREL. 13, Created)  
DT 01-MAY-2000 (TREMUREL. 13, Last sequence update)  
DT 01-MAY-2000 (TREMUREL. 13, Last annotation update)  
DE ELECTRIN 5.  
OS Litorea rubella (Desert tree frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litorea.  
OX NCBI\_TaxID=104895;  
RN [1]

RP SEQUENCE.  
RC TISSUE-SKIN SECRETION;  
RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;  
RT "Peptides from the skin glands of the Australian buzzing tree frog  
RT *Litoria electrica*. Comparison with the skin peptides from *Litoria*  
RT *rubella*.";  
RL Aust. J. Chem. 52:0-0(1999).  
KW Amphibian skin; Amidation.  
FT MOD\_RES 7  
SQ SEQUENCE 7 AA; 834 MW; 6DD05B076DB0B530 CRC64;

Query Match 25.7%; Score 9; DB 13; Length 7;  
Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 YQ 3  
DB 2 YE 3

RESULT 10  
095945  
ID 095945 PRELIMINARY; PRT: 7 AA.  
AC 095945;  
DT 01-FEB-1997 (TREMBLER, 02, Created)  
DT 01-FEB-1997 (TREMBLER, 02, Last sequence update)  
DE 01-DEC-2001 (TREMBLER, 19, Last annotation update)  
DE INSIDE INTRON 5 (FRAGMENT).  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Mitochondrion.  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D273-10B;  
RX MEDLINE=81069885; PubMed=6254986;  
RA Bouliz S.G., Coruzzi G., Thelenfeld B., Tzagoloff A., Macino G.;  
RT "Assembly of the mitochondrial membrane system: Structure and  
RT nucleotide sequence of the gene coding for subunit 1 of yeast  
RT cytochrome oxidase.";  
RL J. Biol. Chem. 255:11927-11941(1980).  
DR EMBL; Y00694; CAA24066.1; -  
KW Mitochondrion.  
FT NON\_TER 1  
SQ SEQUENCE 7 AA; 859 MW; 75B7232362CDC460 CRC64;

Query Match 22.9%; Score 8; DB 8; Length 7;  
Best Local Similarity 25.0%; Pred. No. 5.6e+05;  
Matches 1; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 YOLA 5  
DB 4 WKL5 7

RESULT 11  
09C5B3  
ID 09C5B3 PRELIMINARY; PRT: 7 AA.  
AC 09C5B3;  
DT 01-JUN-2001 (TREMBLER, 17, Created)  
DT 01-JUN-2001 (TREMBLER, 17, Last sequence update)  
DE 01-DEC-2001 (TREMBLER, 19, Last annotation update)  
DE HYPOTHETICAL 0.7 KDA PROTEIN (FRAGMENT).  
GN DIDI 10A-2B.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE=ROOT;  
RX MEDLINE=21171025; PubMed=11277426;  
RA Vercauteren I., Van Der Schueren E., Van Montagu M., Gheysen G.;  
RT "Arabidopsis thaliana genes expressed in the early compatible  
RT interaction with root-knot nematodes.";  
RL Mol. Plant Microbe Interact. 14:288-299(2001).  
DR EMBL; AJ286350; CAB71014.2; -  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 7 AA; 719 MW; 6732C7287EB325D0 CRC64;

Query Match 22.9%; Score 8; DB 10; Length 7;  
Best Local Similarity 66.7%; Pred. No. 5.6e+05;  
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 LAL 6  
DB 5 LVL 7

RESULT 12  
053668  
ID 053668 PRELIMINARY; PRT: 7 AA.  
AC 053668;  
DT 01-NOV-1996 (TREMBLER, 01, Created)  
DT 01-NOV-1996 (TREMBLER, 01, Last sequence update)  
DE 01-DEC-2001 (TREMBLER, 19, Last annotation update)  
DE ORE2 PROTEIN.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRACUE-DANLEY; TISSUE=KIDNEY;  
RX MEDLINE=93396350; PubMed=7667072;  
RA Mandon B., Bellanger A.C., Elalouf J.M.;  
RT "Inverse-PCR-mediated cloning of the promoter for the rat vasopressin  
RT V2 receptor gene.";  
RL Pfugers Arch. 430:12-18(1995).  
DR EMBL; X83264; CAA58237.1; -  
KW SEQUENCE 7 AA; 703 MW; 75A767287DC6D6F0 CRC64;

Query Match 22.9%; Score 8; DB 11; Length 7;  
Best Local Similarity 30.0%; Pred. No. 5.6e+05;  
Matches 2; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 QVAL 6  
DB 2 QAGL 5

RESULT 13  
066113  
ID 066113 PRELIMINARY; PRT: 7 AA.  
AC 066113;  
DT 01-NOV-1996 (TREMBLER, 01, Created)  
DT 01-NOV-1996 (TREMBLER, 01, Last sequence update)  
DE 01-DEC-2001 (TREMBLER, 19, Last annotation update)  
DE C-TERMINUS OF THE VIRAL REPLICASE (FRAGMENT).  
OS Cherry leaf roll virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;  
OC Nepovirus.  
OX NCBI\_TaxID=12615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VALNUT;  
RA Borja M.;  
RL Thesis (1992), Biologia Molecular y Virologia Vegetal, CIT-INIA.

RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WALNUT;  
 RX MEDLINE-96124520; PubMed-8560786;  
 RA Borla M., Sanchez F., Rowhani A., Bruening G., Ponz F.,  
 RT "Long, nearly identical untranslated sequences at the 3' terminal  
 RT regions of the genomic RNAs of cherry leafroll virus (walnut  
 RT strain)."  
 RL Virus genes 10:245-252(1995).  
 DR EMBL: Z34265; CAA84019.1; -.  
 FT NON\_TER  
 SQ SEQUENCE 7 AA; 800 MW; 7417672EBDC6D740 CRC64;

Query Match 22.9%; Score 8; DB 12; Length 7;  
 Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 QIAL 6  
 DB 2 OACL 5

RESULT 14  
 P82100 PRELIMINARY; PRT; 5 AA.  
 AC P82100;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
 DE ELECTRIN 4.  
 OS Litoria rubella (Desert tree frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hyllidae;  
 OC Litoria.  
 OX NCBI\_TaxID-104895;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE-SKIN SECRETION;  
 RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;  
 RT "Peptides from the skin glands of the Australian buzzing tree frog  
 RT Litoria electrica. Comparison with the skin peptides from Litoria  
 RT rubella."  
 RL Aust. J. Chem. 52:0-0(1999).  
 KW Amphibian skin; Amidation.  
 FT MOD\_RES 5  
 SQ SEQUENCE 5 AA; 616 MW; 61F2D1A059A00000 CRC64;

Query Match 20.0%; Score 7; DB 13; Length 5;  
 Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
 Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 LIT 7  
 DB 2 IT 3

RESULT 15  
 P82181 PRELIMINARY; PRT; 6 AA.  
 AC P82181;  
 DT 01-JUN-2000 (TREMblrel. 14, Created)  
 DT 01-JUN-2000 (TREMblrel. 14, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE CHLOROPLAST 50S RIBOSOMAL PROTEIN L10 BETA (FRAGMENT).  
 OS Spinacia oleracea (Spinach).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
 OX NCBI\_TaxID-3562;  
 RN [1]  
 RP SEQUENCE.

RC STRAIN-CV, ALVARO; TISSUE-LEAF;  
 RX MEDLINE-20435798; PubMed-10874046;  
 RA Yamaguchi K., Subramanian A.R.;  
 RT "The plastid ribosomal proteins. Identification of all the proteins in  
 RT the 50 S subunit of an organelle ribosome (chloroplast)."  
 RL J. Biol. Chem. 275:28465-28482(2000)  
 CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES.  
 CC -1- MISCELLANEOUS: ON THE 2D-GEL ITS MW IS: 16.5 KDA.  
 CC -1- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.  
 DR InterPro: IPR001790; Ribosomal\_L10.  
 DR InterPro: IPR002363; Ribosomal\_L10\_eub.  
 DR Pfam: PF00466; Ribosomal\_L10; PARTIAL.  
 DR PROSITE: PS01109; RIBOSOMAL\_L10; PARTIAL.  
 KW Ribosomal protein; Chloroplast; rRNA-binding.  
 FT NON\_TER  
 SQ SEQUENCE 6 AA; 675 MW; 6321B415B05DB000 CRC64;

Query Match 20.0%; Score 7; DB 10; Length 6;  
 Best Local Similarity 33.3%; Pred. No. 5.6e+05;  
 Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 5 ALT 7  
 DB 1 AIS 3

Search completed: July 15, 2002, 13:37:39  
 Job time: 729 sec

10

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:25:25 ; Search time 107.71 Seconds  
(without alignments)  
7.219 Million cell updates/sec

Title: US-09-712-819a-3

Perfect score: 35

Sequence: 1 FVQLALP 7

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 52936

Minimum DB seq length: 0

Maximum DB seq length: 7

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

1: A.Geneseq\_032802:\*

2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*

3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*

4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*

5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*

6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*

7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*

8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*

9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*

10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*

11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*

12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*

13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*

14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*

15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*

16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*

17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*

18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*

19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*

20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*

21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*

22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*

23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57.1	6	19	AAV20207	Human beta-amyloid
2	54.3	5	18	AAW12505	Interleukin-6 anta
3	51.4	4	15	AAW16803	Phylase derived pe
4	51.4	5	18	AAW12496	Interleukin-6 anta
5	51.4	5	18	AAW11532	Interleukin-6 anta
6	51.4	6	13	AAW24623	Immunomodulatory p
7	51.4	6	18	AAW11531	Interleukin-6 anta
8	51.4	6	18	AAW11534	Interleukin-6 anta
9	51.4	6	19	AAW3888	Peptide specific a
10	51.4	6	19	AAW3891	Peptide specific a
11	51.4	6	19	AAW3892	Peptide specific a

Result	Score	Query Match	Length	DB ID	Description
12	51.4	6	19	AAW38872	Peptide specific a
13	51.4	6	19	AAW3781	Peptide #2 used in
14	51.4	6	19	AAW47263	Immunomodulatory p
15	51.4	6	22	AAW88728	Human Interleukin-
16	51.4	7	18	AAW11535	Interleukin-6 anta
17	51.4	7	18	AAW11530	Interleukin-6 anta
18	51.4	7	21	AAW10113	Human angiotensin-
19	51.4	7	22	AAW05447	Synthetic heptapep
20	51.4	7	22	AAW88727	Human Interleukin-
21	51.4	7	22	AAW72813	Antibacterial pept
22	51.4	7	22	AAW12502	Interleukin-6 anta
23	51.4	7	22	AAW12503	Interleukin-6 anta
24	51.4	7	22	AAW12504	Interleukin-6 anta
25	51.4	7	22	AAW12506	Interleukin-6 anta
26	51.4	7	22	AAW12497	Interleukin-6 anta
27	51.4	7	22	AAW11364	Anti-HIV peptide w
28	51.4	7	22	AAW11365	Anti-HIV peptide w
29	51.4	7	22	AAW11366	Anti-HIV peptide w
30	51.4	7	22	AAW11367	Anti-HIV peptide w
31	51.4	7	22	AAW11368	Anti-HIV peptide w
32	51.4	7	22	AAW11369	Anti-HIV peptide w
33	51.4	7	22	AAW11370	Anti-HIV peptide w
34	51.4	7	22	AAW11371	Anti-HIV peptide w
35	51.4	7	22	AAW11372	Anti-HIV peptide w
36	51.4	7	22	AAW11373	Anti-HIV peptide w
37	51.4	7	22	AAW11374	Anti-HIV peptide w
38	51.4	7	22	AAW11375	Anti-HIV peptide w
39	51.4	7	22	AAW11376	Anti-HIV peptide w
40	51.4	7	22	AAW11377	Anti-HIV peptide w
41	51.4	7	22	AAW11378	Anti-HIV peptide w
42	51.4	7	22	AAW11379	Anti-HIV peptide w
43	51.4	7	22	AAW11380	Anti-HIV peptide w
44	51.4	7	22	AAW11381	Anti-HIV peptide w
45	51.4	7	22	AAW11382	Anti-HIV peptide w

#### ALIGNMENTS

RESULT 1

AAV20207 standard; Protein; 6 AA.

AAV20207:

22-0UL-1999 (first entry)

Human beta-amyloid precursor protein mutant fragment 32.

Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer; Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuron; Huntington's disease; multiple sclerosis; alcoholic liver disease; diabetes mellitus type II; microtubule associated protein; Tau; Big Tau; ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M; neurofilament-F; presenilin I; presenilin II; cellular tumor antigen; glial fibrillary acidic protein; GFAP; p53; senaphorin I; HNP-1; bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGp-C; NSP-A; high mobility group protein-C; neuroendocrine specific protein A.

Synthetic.

OS Homo sapiens.

WO9845322-A2.

15-OCT-1998.

02-APR-1998; 98WO-IB00705.

10-APR-1997; 97US-0043163.

(UWU- ) RIJCSUNIV UTRECHT.

(ROTA- ) ROYAL NETHERLANDS ACAD ARTS & SCI.

PA (UYRO-) UNIV ROTTERDAM ERASMUS.  
 PI Burbach JPH, Grosveld FG, Van Leeuwen FW;  
 XX WPI: 1998-609901/51.  
 DR N-PSDB: AAK75753.  
 XX  
 PT Diagnosing disease by detecting frameshift mutations in RNA or  
 PT corresponding protein mutations - used to diagnose cancer and  
 PT neurological diseases, particularly Alzheimer's disease, and also  
 PT for treatment and prevention with specific ribozymes or wild-type  
 PT RNA  
 PS Disclosure: Figure 2; 258bp; English.  
 XX  
 CC This invention describes a novel method for the diagnosis of a disease  
 CC caused by, or associated with, an RNA molecule that has a frameshift  
 CC mutation. The method is used to diagnose age-related diseases, especially  
 CC cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's  
 CC disease, Down's syndrome, myotonic dystrophy, Huntington's disease, II  
 CC multiple sclerosis, alcoholic liver disease, diabetes mellitus type II  
 CC and many others listed) or susceptibility to these disorders. The method  
 CC allows a definitive diagnosis of Alzheimer's disease in living patients,  
 CC at an early stage. It is based on the observation that disease may be  
 CC caused by mutations in RNA rather than DNA. The invention describes the  
 CC use of neuronal system RNA molecules, specifically proteins including  
 CC beta-amyloid precursor protein (beta-APP), the microtubule associated  
 CC proteins tau and big tau, ubiquitin B, apolipoprotein E, microtubule  
 CC associated protein 2 (MAP2), neurofilament-L, neurofilament-M,  
 CC neurofilament-F, presenilin 1, presenilin II, glial fibrillary acidic  
 CC protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma  
 CC 2 (bcl-2) proto-oncogene, semaphorin III, HUP-1, high mobility group  
 CC protein-C (HMGP-C) and neuroendocrine specific protein A.  
 XX  
 SQ Sequence 6 AA;  
 Query Match 57.1%; Score 20; DB 19; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 YQLA 5  
 Db 1111  
 3 yqla 6  
 RESULT 2  
 AAW12505 standard; peptide; 5 AA.  
 ID AAW12505  
 XX  
 AC AAW12505;  
 XX  
 DT 22-APR-1997 (first entry)  
 XX  
 DE Interleukin-6 antagonist 70.  
 XX  
 KW Interleukin-6; IL-6; antagonist; inhibitor; autoimmune disease;  
 KW skin; intestine; systemic lupus erythematosus; chronic rheumatism.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 5 /note="amdated"  
 XX  
 PN JP08311098-A.  
 XX  
 PD 26-NOV-1996.  
 XX  
 PF 22-MAY-1995; 95JP-0146742.  
 XX  
 PR 22-MAY-1995; 95JP-0146742.  
 XX  
 XX 5

PA (DAIL ) DAIICEL CHEM IND LTD.  
 PA (F0J1 ) FUJISHIMA PHARM CO LTD.  
 XX  
 XX WPI: 1997-061811/06.  
 XX  
 PT Interleukin-6 antagonistic peptide(s) comprising arginine - useful  
 PT for treating autoimmune, renal, skin and intestinal diseases  
 XX  
 PS Example 70; Page 12; 20pp; Japanese.  
 XX  
 CC The present peptide is a specific example of new interleukin-6  
 CC antagonists of the general formula X-A-B-D-Y, where X is 1-10 amino  
 CC acids or an amino group protecting group; Y is 1-5 amino acids, a  
 CC carboxyl group protecting group or an amide; A is preferably Arg  
 CC having an opt. protected guanidino group but can be any amino acid;  
 CC D is Arg having an opt. protected guanidino group and B is preferably  
 CC a leu residue but can be any amino acid, including non-natural  
 CC amino acids, opt. having a protected side-chain. The peptides are  
 CC useful for treating autoimmune diseases (e.g. systemic lupus  
 CC erythematosus or chronic rheumatism), renal, skin and intestinal  
 CC diseases.  
 XX  
 SQ Sequence 5 AA;  
 Query Match 54.3%; Score 19; DB 18; Length 5;  
 Best Local Similarity 75.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FYQL 4  
 Db 1111  
 1 fye1 4  
 RESULT 3  
 AAR46803 standard; Protein; 4 AA.  
 ID AAR46803  
 XX  
 AC AAR46803;  
 XX  
 DT 19-ADG-1994 (first entry)  
 XX  
 DE Phytase derived peptide 418(3 phy).  
 XX  
 KW pH 2.5; acid phosphatase; Trichoderma; host: Aspergillus; phytic acid;  
 KW phytate degrading enzyme; PDE; removal; inositol hexaphosphoric acid;  
 KW plant; feed composition; filtration.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9403612-A.  
 XX  
 PD 17-FEB-1994.  
 XX  
 FE 30-JUL-1993; 93WO-FI00310.  
 XX  
 PR 31-JUL-1992; 92US-0923724.  
 XX  
 PA (ALKO-) ALKO LTD.  
 XX  
 PI Cantrell M, Fagerstrom RB, Miettinen-Oinonen ASK;  
 PI Nevalainen HK, Paloheimo MT, Piddington C, Ramosek JA;  
 PI Torkkell TK, Turunen MK;  
 XX  
 DR WPI: 1994-065700/08.  
 XX  
 XX Compns. contg. phytate degrading enzymes - obtd. by expression  
 PT of their genes in Trichoderma, used partic. for producing animal  
 PT feed compns.  
 XX  
 PS Example 4; Page 43; 142pp; English.  
 XX  
 CC The sequences given in AAR46793-824 are peptides derived from the



phytae protein. The phytase protein may be used in the composition of the invention. The DNA encoding the phytase protein may be introduced into a Trichoderma host which then expresses it and the protein is collected from the culture medium. By using Trichoderma as a host for Aspergillus phytate degrading enzymes such as this, a totally different enzyme composition compared to that secreted from Aspergillus results. The enzyme composition can be used for removal of phytic acid or inositol hexaphosphoric acid from raw material, particularly plant material. The composition is used in feed compositions for animals. By using Trichoderma as a source of a composition containing phytate degrading enzymes some difficult downstream processing problems, eg. filtration, that occur with similar Aspergillus compositions are avoided and yields are improved.

Sequence 4 AA:

Query Match 51.4%; Score 18; DB 15; Length 4;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYQ 3  
|||  
Db 1 fYq 3

## RESULT 4

AAW12496 standard; peptide: 5 AA.

AC AAW12496;

DT 22-APR-1997 (first entry)

DE Interleukin-6 antagonist 61.

KW Interleukin-6; IL-6; antagonist; inhibitor; autoimmune disease;  
skin; intestine; systemic lupus erythematosus; chronic rheumatism.

OS Synthetic.

FT Key Location/Qualifiers

FT Modified-site 5 /note= "amidated"

PN JP08311098-A.

PD 26-NOV-1996.

PF 22-MAY-1995; 95JP-0146742.

PR 22-MAY-1995; 95JP-0146742.

PA (DAIL ) DAICEL CHEM IND LTD.

PA (FUJI ) FUJISAWA PHARM CO LTD.

WP1; 1997-061811/06.

Interleukin-6 antagonistic peptide(s) comprising arginine - useful  
for treating autoimmune, renal, skin and intestinal diseases

Example 61; Page 12; 20pp; Japanese.

The present peptide is a specific example of new interleukin-6 antagonists of the general formula X-A-B-D-Y, where X is 1-10 amino acids or an amino group protecting group; Y is 1-5 amino acids, a carboxyl group protecting group or an amide; A is preferably Arg having an opt. protected guanidino group but can be any amino acid; D is Arg having an opt. protected guanidino group and B is preferably a Leu residue but can be any amino acid, including non-natural amino acids, opt. having a protected side-chain. The peptides are useful for treating autoimmune diseases (e.g. systemic lupus erythematosus or chronic rheumatism), renal, skin and intestinal

CC diseases.  
XX  
SQ Sequence 5 AA;

Query Match 51.4%; Score 18; DB 18; Length 5;  
Best Local Similarity 75.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYQL 4  
||:|  
Db 1 fYql 4

## RESULT 5

AAW1532 standard; peptide: 5 AA.

AC AAW1532;

DT 22-APR-1997 (first entry)

DE Interleukin-6 antagonist 5.

KW Interleukin-6; IL-6; antagonist; inhibitor; autoimmune disease;  
skin; intestine; systemic lupus erythematosus; chronic rheumatism.

OS Synthetic.

FT Key Location/Qualifiers

FT Modified-site 5 /note= "optionally amidated"

PN JP08311098-A.

PD 26-NOV-1996.

PF 22-MAY-1995; 95JP-0146742.

PR 22-MAY-1995; 95JP-0146742.

PA (DAIL ) DAICEL CHEM IND LTD.

PA (FUJI ) FUJISAWA PHARM CO LTD.

WP1; 1997-061811/06.

Interleukin-6 antagonistic peptide(s) comprising arginine - useful  
for treating autoimmune, renal, skin and intestinal diseases

Example 5 and Example 6; Page 11; 20pp; Japanese.

The present sequence covers two specific examples of new interleukin-6 antagonists of the general formula X-A-B-D-Y, where X is 1-10 amino acids or an amino group protecting group; Y is 1-5 amino acids, a carboxyl group protecting group or an amide; A is preferably Arg having an opt. protected guanidino group but can be any amino acid; D is Arg having an opt. protected guanidino group and B is preferably a Leu residue but can be any amino acid, including non-natural amino acids, opt. having a protected side-chain. The peptides are useful for treating autoimmune diseases (e.g. systemic lupus erythematosus or chronic rheumatism), renal, skin and intestinal diseases.

Sequence 5 AA;

Query Match 51.4%; Score 18; DB 18; Length 5;  
Best Local Similarity 75.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYQL 4  
||:|  
Db 1 fYql 4

```

RESULT 6
AAW1531
ID AAR24623 standard; peptide; 6 AA.
XX
AC AAR24623;
XX
DT 03-DEC-1992 (first entry)
XX
DE Immunomodulatory peptide.
XX
KW Immunodeficiencies; immunosuppression; T-cell subset; immunotherapy;
KM inflammation; wounds; lymphocyte; vaccine.
XX
OS Synthetic.
XX
PN WO9209628-A.
XX
PD 11-JUN-1992.
XX
PE 22-NOV-1991; 91WO-US08795.
XX
PR 23-NOV-1990; 90US-0617494.
XX
PA (TMMU-) IMMUNODYNAMICS INC.
XX
PI Atkin A;
XX
DR WPI: 1992-217021/26.
XX
PT New synthetic immunomodulatory peptide(s) - for treating
PT immunodeficiencies, immunosuppression and T-cell subset
PT deviations and immuno-therapy of infections, inflammation, wounds
PT etc.
XX
PS Claim 10; Page 36; 52pp; English.
XX
CC The immunomodulatory peptide is a specific example of a peptide cpd.
CC (or an acid or base salt) constructed by combination and/or
CC overlapping of the amino acid sequences A1B1B2A2, A3B3A4B4,
CC B5A5A6B6, B7A7B8B8, A9B9, A10A11, B10A12, and B1B12 (X= Ala, Gly,
CC Ile, Leu, or Val, A1-A12 each- Arg, Asn, Gln, Lys, Phe or Val;
CC B1-B12 each- Asp, Glu, Tyr, Phe or Val. The synthetic peptide may
CC be used for immunomodulation of various immunodeficiencies and
CC immunosuppressed conditions, T-cell subset and lymphocyte deviations,
CC enhancement of a vaccines efficacy, as well as for immunotherapy,
CC including infections, local or systemic complications of non-
CC infectious diseases, postoperative inflammations, wounds and burns.
CC See also AAR24583-R24701.
XX
SQ Sequence 6 AA:

Query Match 51.4%; Score 18; DB 13; Length 6;
Best Local Similarity 75.0%; Pred. No. 6.4e+05;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYOL 4
DB 1 IYI 1
1 IYI 4

RESULT 7
AAW1531
ID AAW1531 standard; peptide; 6 AA.
XX
AC AAW1531;
XX
DT 22-APR-1997 (first entry)
XX
DE Interleukin-6 antagonist 4.
XX

```

```

KM Interleukin-6; IL-6; antagonist; inhibitor; autoimmune disease;
KM skin; intestine; systemic lupus erythematosus; chronic rheumatism.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 6
FT /note= "amidated"
XX
PN JP08311098-A.
XX
PD 26-NOV-1996.
XX
PE 22-MAY-1995; 95JP-0146742.
XX
PR 22-MAY-1995; 95JP-0146742.
XX
PA (DAIL ) DAICEL CHEM IND LTD.
PA (FUJI ) FUJISAWA PHARM CO LTD.
XX
DR WPI: 1997-061811/06.
XX
PT Interleukin-6 antagonistic peptide(s) comprising arginine - useful
PT for treating autoimmune, renal, skin and intestinal diseases
XX
PS Example 1; Page 11; 20pp; Japanese.
XX
CC The present peptide is a specific example of new interleukin-6
CC antagonists of the general formula X-A-B-D-Y, where X is 1-10 amino
CC acids or an amino group protecting group; Y is 1-5 amino acids, a
CC carboxyl group protecting group or an amide; A is preferably Arg;
CC D is Arg having an opt. protected guanidino group but can be any amino acid;
CC A is Leu residue but can be any amino acid, including non-natural
CC amino acids, opt. having a protected side-chain. The peptides are
CC useful for treating autoimmune diseases (e.g. systemic lupus
CC erythematosus or chronic rheumatism), renal, skin and intestinal
CC diseases.
XX
SQ Sequence 6 AA:

Query Match 51.4%; Score 18; DB 18; Length 6;
Best Local Similarity 75.0%; Pred. No. 6.4e+05;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYOL 4
DB 2 IYI 5
2 IYI 5

RESULT 8
AAW1534
ID AAW1534 standard; peptide; 6 AA.
XX
AC AAW1534;
XX
DT 22-APR-1997 (first entry)
XX
DE Interleukin-6 antagonist 9.
XX
KM Interleukin-6; IL-6; antagonist; inhibitor; autoimmune disease;
KM skin; intestine; systemic lupus erythematosus; chronic rheumatism.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 6
FT /note= "amidated"
XX
PN JP08311098-A.
XX
PD 26-NOV-1996.
XX

```

XX 22-MAY-1995; 95JP-0146742.  
 XX 22-MAY-1995; 95JP-0146742.  
 XX (DAIL ) DAICEL CHEM IND LTD.  
 XX (FUJI ) FUJISAMA PHARM CO LTD.  
 XX WPI: 1997-061811/06.  
 XX Interleukin-6 antagonistic peptide(s) comprising arginine - useful  
 PT for treating autoimmune, renal, skin and intestinal diseases  
 XX Example 9; Page 11; 20pp; Japanese.

XX The present peptide is a specific example of new interleukin-6  
 CC antagonists of the general formula X-A-B-D-Y, where X is 1-10 amino  
 CC acids or an amino group protecting group; Y is 1-5 amino acids, a  
 CC carboxyl group protecting group or an amide; A is preferably Arg  
 CC having an opt. protected guanidino group but can be any amino acid;  
 CC D is Arg having an opt. protected guanidino group and B is preferably  
 CC a Leu residue but can be any amino acid, including non-natural  
 CC amino acids, opt. having a protected side-chain. The peptides are  
 CC useful for treating autoimmune diseases (e.g. systemic lupus  
 CC erythematosus or chronic rheumatism), renal, skin and intestinal  
 CC diseases.

SQ Sequence 6 AA:

Query Match 51.4%; Score 18; DB 18; Length 6;  
 Best Local Similarity 75.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYQ 4  
 |||  
 Db 1 fYr1 4

RESULT 9

AAW83888  
 ID AAW83888 standard; peptide; 6 AA.

XX AAW83888;

DT 14-JAN-1999 (first entry)

DE Peptide specific antigen (PSA) cleaved substrate sequence.

KW Human prostate specific antigen; PSA; substrate; prostate cancer;  
 KW metastatic prostate cancer; PSA inhibitor.

OS Synthetic.

PN WO9840738-A1.

PD 17-SEP-1998.

PF 09-MAR-1998; 98MO-US04557.

PR 11-MAR-1997; 97US-0040174.

PA (SLOK ) SLOAN KETTERING INST CANCER RES.

PI Blobel C, Roghani M, Rothman J;

DR WPI: 1998-520833/44.

XX Improved human prostate specific antigen substrates - useful to,  
 PT e.g. measure prostate specific antigen activity; identify  
 PT therapeutically useful inhibitors or diagnose metastatic prostate  
 PT cancer

PS Claim 17; Fig 18; 147pp; English.

XX AAW83856-94 represent peptides cleaved by 50 microgram/ml human prostate  
 CC specific antigen (PSA). They correspond to the consensus sequence  
 CC AAW83855. The PSA substrates are useful in assays which measure PSA  
 CC activity. Such assays may provide a clinical evaluation of patients  
 CC suffering from, or at risk for, prostate cancer. Elevated levels of  
 CC PSA in serum are widely used as a marker of prostate cancer, and the  
 CC substrates can be used e.g. to measure the amount of active PSA in  
 CC patient serum to diagnose, or determine the malignant potential of,  
 CC a patient's prostate disease. They are especially useful to diagnose  
 CC metastatic prostate cancer. The substrates can also be used in assays  
 CC to identify PSA inhibitors, which can be administered to treat or  
 CC prevent prostate cancer.

SQ Sequence 6 AA:

Query Match 51.4%; Score 18; DB 19; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYQ 3  
 |||  
 Db 1 fYq 3

RESULT 10

AAW83891  
 ID AAW83891 standard; peptide; 6 AA.

XX AAW83891;

DT 14-JAN-1999 (first entry)

DE Peptide specific antigen (PSA) cleaved substrate sequence.

KW Human prostate specific antigen; PSA; substrate; prostate cancer;  
 KW metastatic prostate cancer; PSA inhibitor.

OS Synthetic.

PN WO9840738-A1.

PD 17-SEP-1998.

PF 09-MAR-1998; 98MO-US04557.

PR 11-MAR-1997; 97US-0040174.

PA (SLOK ) SLOAN KETTERING INST CANCER RES.

PI Blobel C, Roghani M, Rothman J;

DR WPI: 1998-520833/44.

XX Improved human prostate specific antigen substrates - useful to,  
 PT e.g. measure prostate specific antigen activity; identify  
 PT therapeutically useful inhibitors or diagnose metastatic prostate  
 PT cancer

PS Claim 17; Fig 18; 147pp; English.

XX AAW83856-94 represent peptides cleaved by 50 microgram/ml human prostate  
 CC specific antigen (PSA). They correspond to the consensus sequence  
 CC AAW83855. The PSA substrates are useful in assays which measure PSA  
 CC activity. Such assays may provide a clinical evaluation of patients  
 CC suffering from, or at risk for, prostate cancer. Elevated levels of  
 CC PSA in serum are widely used as a marker of prostate cancer, and the  
 CC substrates can be used e.g. to measure the amount of active PSA in  
 CC patient serum to diagnose, or determine the malignant potential of,  
 CC a patient's prostate disease. They are especially useful to diagnose  
 CC metastatic prostate cancer. The substrates can also be used in assays

CC to identify PSA inhibitors, which can be administered to treat or  
CC prevent prostate cancer.  
XX  
SQ Sequence 6 AA;

Query Match 51.4%; Score 18; DB 19; Length 6;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYO 3  
111  
Db 1 fty 3

## RESULT 11

AAW83892  
ID AAW83892 standard; peptide: 6 AA.

XX AAW83892;

DT 14-JAN-1999 (first entry)

XX Peptide specific antigen (PSA) cleaved substrate sequence.

XX Human prostate specific antigen; PSA; substrate; prostate cancer;  
KW metastatic prostate cancer; PSA inhibitor.

XX Synthetic.

XX WO9840738-A1.

XX 17-SEP-1998.

XX 09-MAR-1998; 98WO-US04557.

XX 11-MAR-1997; 97US-0040174.

PA (SLOK ) SLOAN KETTERING INST CANCER RES.

PI Biobel C, Roghani M, Rothman J;

DR WPI; 1998-520833/44.

XX Improved human prostate specific antigen substrates - useful to,  
PT e.g. measure prostate specific antigen activity, identify  
PT therapeutically useful inhibitors or diagnose metastatic prostate  
PT cancer

PS Claim 17; Fig 18; 147pp; English.

CC AAW83856-94 represent peptides cleaved by 50 microgram/ml human prostate  
CC specific antigen (PSA). They correspond to the consensus sequence  
CC AAW83855. The PSA substrates are useful in assays which measure PSA  
CC activity. Such assays may provide a clinical evaluation of patients  
CC suffering from, or at risk for, prostate cancer. Elevated levels of  
CC PSA in serum are widely used as a marker of prostate cancer, and the  
CC substrates can be used e.g. to measure the amount of active PSA in  
CC a patient's serum to diagnose, or determine the malignant potential of,  
CC a patient's prostate disease. They are especially useful to diagnose  
CC metastatic prostate cancer. The substrates can also be used in assays  
CC to identify PSA inhibitors, which can be administered to treat or  
CC prevent prostate cancer.

XX Sequence 6 AA;

Query Match 51.4%; Score 18; DB 19; Length 6;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYO 3  
111;

Db 1 fty 3

## RESULT 12

AAW83872  
ID AAW83872 standard; peptide: 6 AA.

XX AAW83872;

DT 14-JAN-1999 (first entry)

XX Peptide specific antigen (PSA) cleaved substrate sequence.

XX Human prostate specific antigen; PSA; substrate; prostate cancer;  
KW metastatic prostate cancer; PSA inhibitor.

XX Synthetic.

XX WO9840738-A1.

XX 17-SEP-1998.

XX 09-MAR-1998; 98WO-US04557.

XX 11-MAR-1997; 97US-0040174.

PA (SLOK ) SLOAN KETTERING INST CANCER RES.

PI Biobel C, Roghani M, Rothman J;

DR WPI; 1998-520833/44.

XX Improved human prostate specific antigen substrates - useful to,  
PT e.g. measure prostate specific antigen activity, identify  
PT therapeutically useful inhibitors or diagnose metastatic prostate  
PT cancer

PS Claim 17; Fig 18; 147pp; English.

CC AAW83856-94 represent peptides cleaved by 50 microgram/ml human prostate  
CC specific antigen (PSA). They correspond to the consensus sequence  
CC AAW83855. The PSA substrates are useful in assays which measure PSA  
CC activity. Such assays may provide a clinical evaluation of patients  
CC suffering from, or at risk for, prostate cancer. Elevated levels of  
CC PSA in serum are widely used as a marker of prostate cancer, and the  
CC substrates can be used e.g. to measure the amount of active PSA in  
CC a patient's serum to diagnose, or determine the malignant potential of,  
CC a patient's prostate disease. They are especially useful to diagnose  
CC metastatic prostate cancer. The substrates can also be used in assays  
CC to identify PSA inhibitors, which can be administered to treat or  
CC prevent prostate cancer.

XX Sequence 6 AA;

Query Match 51.4%; Score 18; DB 19; Length 6;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYO 3  
111  
Db 1 fty 3

## RESULT 13

AAW33781  
ID AAW33781 standard; peptide: 6 AA.

XX AAW33781;

DT 19-JUN-1998 (first entry)

XX Peptide #2 used in immunomodulating dimer peptide.



PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Saxinger C;

XX WPI; 2001-244395/25.

DR  
XX Polypeptides which inhibit the binding of interleukin (IL)-6 ligand  
PT with the IL-6 receptor, and the nucleic acids that encode them, useful  
PT for treating e.g. inflammation and autoimmune diseases -

XX  
PS Example 3; Page 46; 98pp; English.

XX  
CC The present invention describes a number of peptides which are able to  
CC bind to the human interleukin-6 (IL-6) receptor and inhibit IL-6 binding.  
CC These are useful in the treatment of diseases associated with abnormal  
CC IL-6 expression, including multiple myeloma, plasmacytoma, haematological  
CC diseases such as plasma cell dyscrasias, leukaemia and lymphoma,  
CC mesangial proliferative glomerulonephritis, polyclonal B cell activation  
CC conditions, allergies, rheumatoid arthritis, diabetes, multiple  
CC sclerosis, septic shock, infections, post-menopausal osteoporosis,  
CC chronic immune deficiency, autoimmune diseases and inflammatory diseases.

XX  
SQ Sequence 6 AA;

Query Match 51.4%; Score 18; DB 22; Length 6;

Best Local Similarity 75.0%; Pred. No. 6.4e+05; Mismatches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 F10L 4

Db 3 tyr1 6

Search completed: July 15, 2002, 13:25:26  
Job time: 1457 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 15, 2002, 13:26:13 ; Search time 40.89 seconds  
(without alignments)  
4.181 Million cell updates/sec

Title: US-09-712-819A-3  
Perfect score: 35  
Sequence: 1 FYQLALT 7

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 39160

Minimum DB seq length: 0  
Maximum DB seq length: 7

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/prodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/prodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/prodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/prodata/2/1aa/PCYUS.COMB.pep.\*  
6: /cgn2\_6/prodata/2/1aa/Backfilest.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	60.0	7	4	US-09-183-217-7
2	18	51.4	4	1	US-07-923-724-30
3	18	51.4	4	2	US-08-609-426A-30
4	18	51.4	4	2	US-08-374-652C-22
5	17	48.6	7	1	US-08-350-884-30
6	17	48.6	7	1	US-08-709-173-30
7	17	48.6	7	2	US-08-709-173-30
8	17	48.6	7	3	US-08-946-329A-26
9	17	48.6	7	4	US-08-567-357A-26
10	17	48.6	7	4	US-08-729-743A-26
11	17	48.6	7	4	US-09-457-046B-40
12	16	45.7	5	1	US-08-238-089-23
13	16	45.7	5	1	US-08-238-089-23
14	16	45.7	5	1	US-08-432-651A-23
15	16	45.7	5	1	US-08-432-651A-23
16	16	45.7	5	3	US-08-962-962C-23
17	16	45.7	5	3	US-08-962-962C-23
18	16	45.7	5	3	US-08-962-962C-23
19	16	45.7	5	5	PCT-US95-05560-25
20	16	45.7	5	5	PCT-US95-05560-25
21	16	45.7	6	5	PCT-US93-01669-50
22	16	45.7	7	1	US-08-193-977-15
23	16	45.7	7	1	US-08-354-240A-8
24	16	45.7	7	4	US-09-220-081-10
25	16	45.7	7	4	US-09-258-754-271
26	16	45.7	7	4	US-09-042-107-271
27	16	45.7	7	6	US-09-457-046B-39

28	15	42.9	4	6	5227466-26	Patent No. 5227466
29	15	42.9	5	1	US-08-392-646-12	Sequence 12, Appl
30	15	42.9	5	1	US-08-392-646-13	Sequence 13, Appl
31	15	42.9	5	2	US-08-612-785B-21	Sequence 21, Appl
32	15	42.9	5	3	US-08-469-141A-40	Sequence 40, Appl
33	15	42.9	5	4	US-08-703-675C-34	Sequence 34, Appl
34	15	42.9	5	4	US-08-617-267C-21	Sequence 21, Appl
35	15	42.9	5	4	US-08-617-267C-40	Sequence 40, Appl
36	15	42.9	5	5	PCT-US95-13794-40	Sequence 40, Appl
37	15	42.9	6	1	US-08-106-493A-3	Sequence 3, Appl1
38	15	42.9	6	1	US-08-429-264-3	Sequence 3, Appl1
39	15	42.9	6	1	US-08-374-843B-27	Sequence 27, Appl
40	15	42.9	6	2	US-08-317-310A-54	Sequence 54, Appl
41	15	42.9	6	2	US-08-905-420-27	Sequence 27, Appl
42	15	42.9	6	2	US-08-540-412-114	Sequence 114, App
43	15	42.9	6	2	US-08-540-412-115	Sequence 115, App
44	15	42.9	6	2	US-08-540-412-135	Sequence 135, App
45	15	42.9	6	2	US-08-540-412-159	Sequence 159, App

## ALIGNMENTS

```
RESULT 1
US-09-183-217-7
: Sequence 7, Application US/09183217A
: Patent No. 6153194
: GENERAL INFORMATION:
: APPLICANT: Skare, Jonathan T.
: APPLICANT: Shang, Ellen S.
: APPLICANT: Champion, Cheryl I.
: APPLICANT: Bianco, David R.
: APPLICANT: Miller, James N.
: APPLICANT: Lovett, Michael A.
: APPLICANT: Mizabekov, Tajib A.
: APPLICANT: Kagan, Bruce L.
: APPLICANT: Tempst, Paul
: APPLICANT: Foley, Denise M.
: TITLE OR INVENTION: BORRELLIA BURGDOFFERI OUTER MEMBRANE PROTEINS
: FILE REFERENCE: UC Case No. 6153194 96-059-3/Skare et al.
: CURRENT APPLICATION NUMBER: US/09/183,217A
: PRIOR FILING DATE: 1998-10-29
: PRIOR FILING DATE: 1997-01-22
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 7
: LENGTH: 7
: TYPE: PRT
: ORGANISM: Borrelia burgdorferi
US-09-183-217-7

Query Match 60.0%; Score 21; DB 4; Length 7;
Best local similarity 66.7%; Pred. No. 1.7e+05;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 VOLALT 7
Db 1 YKLGLT 6

RESULT 2
US-07-923-724-30
: Sequence 30, Application US/07923724
: Patent No. 5780292
: GENERAL INFORMATION:
: APPLICANT: Nevalainen, Helena K.M.
: APPLICANT: Palohelmo, Maria T.
: APPLICANT: Miettinen-Oinonen, Arja S.K.
: APPLICANT: Torkkeli, Tuula K.
: APPLICANT: Cantrell, Michael
: APPLICANT: Piddington, Christopher S.
```

```

? APPLICANT: Rambosek, John A.
? APPLICANT: Turunen, Marja K.
? APPLICANT: Fagerstr m, Richard B.
? TITLE OF INVENTION: Production of Phytase Degrading Enzymes
? TITLE OF INVENTION: In Trichoderma
? NUMBER OF SEQUENCES: 66
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Sterne, Kessler, Goldstein & Fox
? STREET: 1100 New York Avenue, Suite 600
? CITY: Washington
? STATE: D.C.
? COUNTRY: U.S.A.
? ZIP: 20005
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/07/923,724
? FILING DATE: 31-JUL-1992
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/496,155
? FILING DATE: 19-MAR-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/044,077
? FILING DATE: 29-APR-1987
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: UK 8610600
? FILING DATE: 30-APR-1986
? ATTORNEY/AGENT INFORMATION:
? NAME: Cimbalg, Michele A.
? REGISTRATION NUMBER: 33,851
? REFERENCE/DOCKET NUMBER: 1050.0240004
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 371-2540
? TELEFAX: (202) 371-2540
? INFORMATION FOR SEQ ID NO: 30:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4 amino acids
? TYPE: amino acid
? TOPOLOGY: both
?
? US-07-923-724-30
?
? Query Match 51.4%; Score 18; DB 1; Length 4;
? Best Local Similarity 100.0%; Pred. No. 1.7e+05;
? Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
?
? QY 1 FYQ 3
? 111
? DB 1 FYQ 3
?
? RESULT 3
? US-08-609-426A-30
? Sequence 30, Application US/08609426A
? Patent No. 5830733
? GENERAL INFORMATION:
? APPLICANT: Nevalainen, Helena K.M.
? APPLICANT: Paloheimo, Marja T.
? APPLICANT: Miettinen-Oinonen, Arja S.K.
? APPLICANT: Torkkeli, Tuula K.
? APPLICANT: Cantrell, Michael
? APPLICANT: Piddington, Christopher S.
? APPLICANT: Rambosek, John A.
? APPLICANT: Turunen, Marja K.
? APPLICANT: Fagerstr m, Richard B.
? APPLICANT: Houston, Christine S.
? TITLE OF INVENTION: Production of Phytase Degrading Enzymes
? TITLE OF INVENTION: In Trichoderma
? NUMBER OF SEQUENCES: 69

```

```

? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Sterne, Kessler, Goldstein & Fox
? STREET: 1100 New York Avenue, Suite 600
? CITY: Washington
? STATE: D.C.
? COUNTRY: U.S.A.
? ZIP: 20005
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/609,426A
? FILING DATE: 01-MAR-1996
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/923,724
? FILING DATE: 31-JUL-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/496,155
? FILING DATE: 19-MAR-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/044,077
? FILING DATE: 29-APR-1987
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: UK 8610600
? FILING DATE: 30-APR-1986
? ATTORNEY/AGENT INFORMATION:
? NAME: Reed, Grant E.
? REGISTRATION NUMBER: P-41,264
? REFERENCE/DOCKET NUMBER: 1050.0080001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 371-2600
? TELEFAX: (202) 371-2540
? INFORMATION FOR SEQ ID NO: 30:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4 amino acids
? TYPE: amino acid
? TOPOLOGY: both
?
? US-08-609-426A-30
?
? Query Match 51.4%; Score 18; DB 2; Length 4;
? Best Local Similarity 100.0%; Pred. No. 1.7e+05;
? Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
?
? QY 1 FYQ 3
? 111
? DB 1 FYQ 3
?
? RESULT 4
? US-08-374-652C-22
? Sequence 22, Application US/08374652C
? Patent No. 5834286
? GENERAL INFORMATION:
? APPLICANT: NEVALAINEN, HELENA K.M.
? APPLICANT: PALOHEIMO, MARJA T.
? APPLICANT: FAGERSTROM, RICHARD B.
? APPLICANT: MIETTINEN-OINONEN, ARJA S.
? APPLICANT: TURUNEN, MARJA K.
? APPLICANT: RAMBOSEK, JOHN A.
? APPLICANT: PIDDINGTON, CHRISTOPHER S.
? APPLICANT: HOUSTON, CHRISTINE S.
? APPLICANT: CANTRELL, MICHAEL A.
? TITLE OF INVENTION: RECOMBINANT CELLS, DNA CONSTRUCTS,
? TITLE OF INVENTION: VECTORS AND METHODS FOR EXPRESSING PHYTASE DEGRADING
? TITLE OF INVENTION: ENZYMES IN DESIRED RATIOS
? NUMBER OF SEQUENCES: 94
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
? STREET: 1100 NEW YORK AVENUE, SUITE 600

```



CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/374,652C  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/07058  
FILING DATE: 27-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/925,401  
FILING DATE: 31-JUL-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: REED, GRANT E.  
REGISTRATION NUMBER: 41,264  
REFERENCE/DOCKET NUMBER: 1050.071001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
US-08-374-652C-22

Query Match 51.4%; Score 18; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYO 3  
111  
Db 1 FYO 3

RESULT 5  
US-08-350-884-30  
Sequence 30, Application US/08350884  
Patent No. 5585258  
GENERAL INFORMATION:  
APPLICANT: HOUGHTON, MICHAEL  
APPLICANT: CHOO, QUI LIM  
TITLE OF INVENTION: HEPATITIS C VIRUS PROTEASE  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/350,884  
FILING DATE: 06-DEC-1994  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/680,296  
FILING DATE: 04-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: CIOTTI, THOMAS E.  
REGISTRATION NUMBER: 21,013  
REFERENCE/DOCKET NUMBER: 22300-20100.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-350-884-30

Query Match 48.6%; Score 17; DB 1; Length 7;  
Best Local Similarity 60.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 YDIAL 6  
111  
Db 2 YDIAL 6

RESULT 6  
US-08-709-173-30  
Sequence 30, Application US/08709173  
Patent No. 5712145  
GENERAL INFORMATION:  
APPLICANT: HOUGHTON, MICHAEL  
APPLICANT: CHOO, QUI LIM  
TITLE OF INVENTION: HEPATITIS C VIRUS PROTEASE  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/709,173  
FILING DATE: 06-SEP-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/680,296  
FILING DATE: 04-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: CIOTTI, THOMAS E.  
REGISTRATION NUMBER: 21,013  
REFERENCE/DOCKET NUMBER: 22300-20100.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-709-173-30

Query Match 48.6%; Score 17; DB 1; Length 7;  
Best Local Similarity 60.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 YOLAL 6  
1 : 1 : 1 :  
Db 2 YDIAL 6

## RESULT 7

US-08-709-177-30  
Sequence 30, Application US/08709177

Patent No. 5885799

GENERAL INFORMATION:

APPLICANT: HOUGHTON, MICHAEL

APPLICANT: CHOO, QUI LIM

TITLE OF INVENTION: HEPATITIS C VIRUS PROTEASE

NUMBER OF SEQUENCES: 86

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FORSTER

STREET: 755 Page Mill Road

CITY: Palo Alto

STATE: California

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/709,177

FILING DATE: 06-SEP-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/660,296

FILING DATE: 04-APR-1991

ATTORNEY/AGENT INFORMATION:

NAME: CIOTTI, THOMAS E.

REGISTRATION NUMBER: 21,013

REFERENCE/DOCKET NUMBER: 22300-20100.20

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 813-5600

TELEFAX: (415) 494-0792

TELEX: 706141

INFORMATION FOR SEQ ID NO: 30:

SEQUENCE CHARACTERISTICS:

LENGTH: 7 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-709-177-30

Query Match 48.6%; Score 17; DB 2; Length 7;

Best Local Similarity 60.0%; Pred. No. 1.7e+05;

Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 YOLAL 6  
1 : 1 : 1 :  
Db 2 YDIAL 6

RESULT 8

US-08-946-329A-26

Sequence 26, Application US/08946329A

Patent No. 6057091

GENERAL INFORMATION:

APPLICANT: Beachy, Philip A.

APPLICANT: Porter, Jeffrey A.

TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES

NUMBER OF SEQUENCES: 109

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: FASTSEQ for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/946,329A

FILING DATE: 07-OCT-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/061,323

FILING DATE: 07-OCT-1996

APPLICATION NUMBER: 08/729,743

FILING DATE: 10-JUL-1996

APPLICATION NUMBER: 08/567,357

FILING DATE: 04-DEC-1995

APPLICATION NUMBER: 08/349,498

FILING DATE: 02-DEC-1994

ATTORNEY/AGENT INFORMATION:

NAME: Hallie, Lisa A.

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 07265/140001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619/678-5070

TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 7 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-946-329A-26

Query Match 48.6%; Score 17; DB 3; Length 7;

Best Local Similarity 100.0%; Pred. No. 1.7e+05;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 LAIT 7  
1 : 1 : 1 :  
Db 1 LAIT 4

RESULT 9

US-08-567-357A-26

Sequence 26, Application US/08567357A

Patent No. 6132728

GENERAL INFORMATION:

APPLICANT: Beachy, Philip A.

APPLICANT: Moon, Randall T.

TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: FASTSEQ for Windows Version 2.0b

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/567,357A  
FILING DATE: 04-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/349,498  
FILING DATE: 02-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/080001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-567-357A-26

Query Match 48.6%; Score 17; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 LALT 7  
||||  
Db 1 LALT 4

RESULT 10  
US-08-729-743A-26  
Sequence 26, Application US/08729743A  
Patent No. 6214794  
GENERAL INFORMATION:  
APPLICANT: Beachy, Phillip A.  
APPLICANT: Moon, Randall T.  
APPLICANT: Porter, Jeffrey A.  
TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/729,743A  
FILING DATE: 07-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/567,357  
FILING DATE: 04-DEC-1995  
APPLICATION NUMBER: 08/349,498  
FILING DATE: 02-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/099001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
US-08-729-743A-26

Query Match 48.6%; Score 17; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 LALT 7  
||||  
Db 1 LALT 4

RESULT 11  
US-09-457-046B-40  
Sequence 40, Application US/09457046B  
Patent No. 6287835  
GENERAL INFORMATION:  
APPLICANT: Croteau, Rodney et al.  
TITLE OF INVENTION: Transacylases of the Pacifitaxel Biosynthetic Pathway  
FILE REFERENCE: 53679  
CURRENT APPLICATION NUMBER: US/09/457,046B  
CURRENT FILING DATE: 1999-12-07  
NUMBER OF SEQ ID NOS: 74  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 40  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:consensus  
US-09-457-046B-40

Query Match 48.6%; Score 17; DB 4; Length 7;  
Best Local Similarity 60.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYOLA 5  
||||  
Db 1 FYOLA 5

RESULT 12  
US-08-238-089-23  
Sequence 23, Application US/08238089  
Patent No. 5589460  
GENERAL INFORMATION:  
APPLICANT: Abajian, Henry B.  
APPLICANT: No. 35894601e, John F.  
APPLICANT: Hlavka, Joseph J.  
TITLE OF INVENTION: TRI-, TETRA-, PENTA-, AND POLYPEPTIDES  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lewis F. Gould, Jr.  
STREET: 1700 Market Street, Suite 3232  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/238,089  
FILING DATE: 04-MAY-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Gould, Jr., Lewis F.

REGISTRATION NUMBER: 25,057  
REFERENCE/DOCKET NUMBER: 3297-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 575-6020  
TELEFAX: (215) 575-6015  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 5  
OTHER INFORMATION: /label= Gly-NH2  
OTHER INFORMATION: /note= "A modified Gly residue: an amine group  
OTHER INFORMATION: replaces a hydroxyl group at the carboxy terminus."  
US-08-238-089-23

Query Match 45.7%; Score 16; DB 1; Length 5;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYOL 4  
DB 1 FYPL 4

RESULT 13  
US-08-238-089-25  
Sequence 25, Application US/08238089  
Patent No. 5589460  
GENERAL INFORMATION:  
APPLICANT: Abajian, Henry B.  
APPLICANT: No. 5589460le, John F.  
APPLICANT: Hlavka, Joseph J.  
TITLE OF INVENTION: TRI-, TETRA-, PENTA-, AND POLYPEPTIDES  
TITLE OF INVENTION: AND THEIR THERAPEUTIC USE AS AN ANTIDEPRESSANT AGENT  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lewis F. Gould, Jr.  
STREET: 1700 Market Street, Suite 3232  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/238, 089  
FILING DATE: 04-MAY-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Gould, Jr., Lewis F.  
REGISTRATION NUMBER: 25,057  
REFERENCE/DOCKET NUMBER: 3297-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 575-6020  
TELEFAX: (215) 575-6015  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site

LOCATION: 3  
OTHER INFORMATION: /label= 4Hyp  
OTHER INFORMATION: /note= "Amino acid #3 is either cis- or trans- 4Hyp"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 5  
OTHER INFORMATION: /label= Gly-NH2  
OTHER INFORMATION: /note= "A modified Gly residue: an amine group  
OTHER INFORMATION: replaces a hydroxyl group at the carboxy terminus."  
US-08-238-089-25

Query Match 45.7%; Score 16; DB 1; Length 5;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYOL 4  
DB 1 FYXL 4

RESULT 14  
US-08-432-651A-23  
Sequence 23, Application US/08432651A  
Patent No. 5767083  
GENERAL INFORMATION:  
APPLICANT: Abajian, Henry B.  
APPLICANT: No. 5767083le, John F.  
APPLICANT: Hlavka, Joseph J.  
TITLE OF INVENTION: TRI-, TETRA-, PENTA-, AND POLYPEPTIDES  
TITLE OF INVENTION: AND THEIR THERAPEUTIC USE AS AN ANTIDEPRESSANT AGENT  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lewis F. Gould, Jr.  
STREET: 1700 Market Street, Suite 3232  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/432,651A  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Gould, Jr., Lewis F.  
REGISTRATION NUMBER: 25,057  
REFERENCE/DOCKET NUMBER: 3297-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 575-6020  
TELEFAX: (215) 575-6015  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 5  
OTHER INFORMATION: /label= Gly-NH2  
OTHER INFORMATION: /note= "A modified Gly residue: an amine group  
OTHER INFORMATION: replaces a hydroxyl group at the carboxy terminus."  
US-08-432-651A-23

Query Match 45.7%; Score 16; DB 1; Length 5;

Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYOL 4  
Db 1 FYPL 4

Search completed: July 15, 2002, 13:26:14  
Job time: 494 sec

RESULT 15

US-08-432-651A-25  
; Sequence 25, Application US/08432651A

; Patent No. 5767083

GENERAL INFORMATION:

APPLICANT: Adajian, Henry B.  
APPLICANT: No. 5767083le, John F.

APPLICANT: Hlavka, Joseph J.  
TITLE OF INVENTION: TRI-, TETRA-, PENTA-, AND POLYPEPTIDES

TITLE OF INVENTION: AND THEIR THERAPEUTIC USE AS AN ANTIDEPRESSANT AGENT  
NUMBER OF SEQUENCES: 73

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lewis F. Gould, Jr.

STREET: 1700 Market Street, Suite 3232  
City: Philadelphia

STATE: PA  
COUNTRY: US

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/432,651A

FILING DATE:  
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:  
NAME: Gould, J.F., Lewis F.

REGISTRATION NUMBER: 25,057  
REFERENCE/DOCKET NUMBER: 3297-5

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 575-6020

TELEFAX: (215) 575-6015  
INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids

TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: peptide

FEATURE:  
NAME/KEY: Modified-site

LOCATION: 3  
OTHER INFORMATION: /label-4Hyp

OTHER INFORMATION: /note- "amino acid #3 is either cis- or trans-"  
OTHER INFORMATION: 4Hyp

FEATURE:  
NAME/KEY: Modified-site

LOCATION: 5  
OTHER INFORMATION: /label-Gly-NH2

OTHER INFORMATION: /note- "A modified Gly residue: an amine group  
OTHER INFORMATION: replaces a hydroxyl group at the carboxy

OTHER INFORMATION: terminus."  
US-08-432-651A-25

Query Match 45.7%; Score 16; DB 1; Length 5;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYOL 4  
Db 1 FYXL 4

10

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:35:10 ; Search time 14.12 Seconds  
(without alignments)  
54,442 Million cell updates/sec

Title: US-09-712-819A-4

Perfect score: 46

Sequence: 1 PKLLIYMA 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 28338 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 603

Minimum DB seq length: 0  
Maximum DB seq length: 8

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

PIR-71:\*  
1: pirl:\*  
2: pirl:\*  
3: pirl:\*  
4: pirl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	39.1	7	2	glucuronosyltransf
2	17	37.0	5	2	copper resistance
3	16	34.8	7	2	hypothetical prote
4	16	34.8	8	2	thymic humoral fac
5	15	32.6	5	2	bradykinin-potenti
6	15	32.6	6	2	pev-kinin 2 - pena
7	15	32.6	7	2	H2 class I protein
8	15	32.6	7	2	glycoprotein compo
9	14	30.4	5	2	gut pentapeptide -
10	14	30.4	6	2	cerebellar degener
11	14	30.4	7	2	dermorphin (Trp-4,
12	14	30.4	7	2	tubulin beta-3 cha
13	14	30.4	8	2	chitin-binding pro
14	13	28.3	6	2	cytotoxic T-lympho
15	13	28.3	7	2	S33246
16	13	28.3	8	2	neuroendulatory pe
17	13	28.3	8	2	angiotensin-conver
18	13	28.3	8	2	glycine reductase
19	13	28.3	8	2	hypothetical prote
20	13	28.3	8	2	leucoproteinkin - M
21	12	26.1	5	2	blood cell protein
22	12	26.1	6	2	Ig heavy chain CRD
23	12	26.1	6	2	alcohol dehydrogen
24	12	26.1	6	2	repetitive protein
25	12	26.1	6	2	sperm acrosomal pr
26	12	26.1	7	2	dermorphin (lys-7)
27	12	26.1	7	2	platelet aggregati
28	12	26.1	7	2	neuroendulatory pe
29	12	26.1	8	2	major fat-globule
					P element, P cytot

30	12	26.1	8	2	A14683	aspartate transami
31	12	26.1	8	2	A61467	penicillin - Adeli
32	11	23.9	3	3	F37196	bradykinin-potenti
33	11	23.9	4	2	A34626	RPCH-related neuro
34	11	23.9	4	2	B53284	T-cell receptor be
35	11	23.9	4	2	PT0661	T-cell receptor be
36	11	23.9	5	2	A32516	cholecystokinin-5
37	11	23.9	5	2	A60803	neuropeptide - sea
38	11	23.9	5	2	PT0308	Ig heavy chain CRD
39	11	23.9	5	2	PT0729	T-cell receptor be
40	11	23.9	5	2	PT0580	T-cell receptor be
41	11	23.9	6	2	B34835	dnaa protein - pse
42	11	23.9	6	2	A31263	dihydrofolate redu
43	11	23.9	6	2	B31263	dihydrofolate redu
44	11	23.9	6	2	A61068	locustakinin - mig
45	11	23.9	6	2	PT0629	T-cell receptor be

## ALIGNMENTS

RESULT 1  
PX0008  
glucuronosyltransferase (EC 2.4.1.17), hepatic - rat (fragment)  
N:Alternate names: UDP-glucuronosyltransferase  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 17-Jul-1992 #sequence,revision 17-Jul-1992 #text\_change 07-Feb-1997  
C:Accession: PX0008  
R:Okada, H.; Yuasa, A.; Sato, R.  
U. Biochem. 104, 531-536, 1988  
A:Title: Purification and properties of a form of UDP-glucuronosyltransferase from liver  
A:Reference number: PX0008; MUID:89197852  
A:Accession: PX0008  
A:Molecule type: protein  
A:Residues: 1-7 <YOK>  
C:Keywords: glycosyltransferase; hexosyltransferase; liver

Query Match 39.1% Score 18; DB 2; Length 7;  
Best Local Similarity 60.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 KLLY 6  
| | | | |  
Db 2 KLLVM 6

RESULT 2  
C41225  
copper resistance protein - Pseudomonas syringae pv. tomato (fragment)  
C:Species: Pseudomonas syringae pv. tomato  
C>Date: 19-Jun-1992 #sequence,revision 19-Jun-1992 #text\_change 24-Jun-1993  
C:Accession: C41225  
R:Chen, J.S.; Cooksey, D.A.  
Proc. Natl. Acad. Sci. U.S.A. 88, 8915-8919, 1991  
A:Title: Copper resistance in Pseudomonas syringae mediated by periplasmic and outer  
A:Reference number: A41225; MUID:92020961  
A:Accession: C41225  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-5 <CHA>

Query Match 37.0% Score 17; DB 2; Length 5;  
Best Local Similarity 75.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLL 4  
| | | |  
Db 2 PKLLV 5

RESULT 3

S09652  
hypothetical protein (aacC2 3' region) - Enterobacter cloacae (fragment)  
C:Species: Enterobacter cloacae  
C>Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 08-Oct-1999  
C:Accession: S09652  
R:Vlieghehart, J.S.; Ketelaar-van Gaalen, P.A.G.; van de Klundert, J.A.M.  
Antilebdo. Agents Chemother. 33, 1153-1159, 1989  
A:Title: Nucleotide sequence of the aacC2 gene, a gentamicin resistance determinant invg  
A:Reference number: S09651; MUID:90024972  
A:Accession: S09652  
A:Molecule type: DNA  
A:Residues: 1-7 <VLI>  
A:Cross-references: EMBL:X51534; NID:g40876; PIDN:CA35914.1; PID:g581034

Query Match 34.8%; Score 16; DB 2; Length 7;  
Best Local Similarity 30.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 LIW 7  
1 1  
Db 1 MITW 4

RESULT 4  
A28719  
thymic humoral factor gamma-2 - bovine (fragment)  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 18-Jun-1993  
C:Accession: A28719  
R:Bursstein, Y.; Buchner, V.; Pecht, M.; Trainin, N.  
Biochemistry 27, 4066-4071, 1988  
A:Title: Thymic humoral factor gamma-2: purification and amino acid sequence of an immun  
A:Reference number: A28719; MUID:88326920  
A:Accession: A28719  
A:Molecule type: protein  
A:Residues: 1-8 <BUR>

Query Match 34.8%; Score 16; DB 2; Length 8;  
Best Local Similarity 75.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKL 4  
1 1  
Db 5 PKFL 8

RESULT 5  
G37196  
bradykinin-potentiating peptide 7 - island jararaca  
C:Species: Bothrops insularis (Island Jararaca)  
C>Date: 14-Feb-1992 #sequence\_revision 01-Dec-1992 #text\_change 05-Aug-1994  
C:Accession: G37196  
R:Cintrá, A.C.O.; Vieira, C.A.; Giglio, J.R.  
J. Protein Chem. 9, 221-227, 1990  
A:Title: Primary structure and biological activity of bradykinin potentiating peptides f  
A:Reference number: A37196; MUID:90351557  
A:Accession: G37196  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-5 <CIN>  
A:Keywords: pyroglutamic acid  
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 32.6%; Score 15; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 WA 8  
1 1  
Db 3 WA 4

RESULT 6  
PD0028  
pav-kinin 2 - penaeid shrimp (Penaeus vannamei) (fragment)  
C:Species: Penaeus vannamei  
C>Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 19-May-2000  
C:Accession: PD0028  
R:Nieto, J.; Veeleart, D.; Derna, R.; Maelkens, E.; Cerslaens, A.; Coast, G.; Devree  
Biochem. Biophys. Res. Commun. 248, 406-411, 1998  
A:Title: Identification of one tachykinin- and two kinin-related peptides in the bral  
A:Reference number: PD0027; MUID:98342103  
A:Accession: PD0028  
A:Molecule type: protein  
A:Residues: 1-6 <NIE>  
A:Comment: This peptide belongs to myotropic neuropeptides.

Query Match 32.6%; Score 15; DB 2; Length 6;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 WA 8  
1 1  
Db 5 WA 6

RESULT 7  
P00727  
H2 class I protein [imported] - rice (fragment)  
C:Species: Oryza sativa (rice)  
C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: P00727  
R:Komatsu, S.; Kajiwara, H.; Hirano, H.  
Theor. Appl. Genet. 86, 935-942, 1993  
A:Title: A rice protein library: a data-file of rice proteins separated by two-dimens  
A:Reference number: P00696  
A:Accession: P00727  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-7 <KOM>

Query Match 32.6%; Score 15; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 WA 8  
1 1  
Db 1 WA 2

RESULT 8  
E48394  
glycoprotein component 16/major fat-globule membrane protein/MFG-E8 homolog - bovine  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 19-Nov-1993 #sequence\_revision 18-Nov-1994 #text\_change 07-Feb-1997  
C:Accession: E48394  
R:Mathier, I.H.; Banghart, L.R.; Lane, W.S.  
Biochem. Mol. Biol. Int. 29, 545-554, 1993  
A:Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-  
II-like sequences.  
A:Reference number: A48394; MUID:93250576  
A:Accession: E48394  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-7 <MAT>  
A:Experimental source: milk  
A>Note: sequence extracted from NCBI backbone (NCBIF:131450)  
C:Keywords: glycoprotein

Query Match 32.6%; Score 15; DB 2; Length 7;



Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 WA 8  
1 1  
Db 1 WA 2

## RESULT 9

JH0253  
gut pentapeptide - Japanese eel  
C:Species: *Anguilla japonica* (Japanese eel)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 11-Apr-1995  
C:Accession: JH0253  
R:Usaka, T.; Ikeda, T.; Kubota, I.; Muneoka, Y.; Ando, M.  
Biochem. Biophys. Res. Commun. 180: 828-832, 1991  
A:Title: Structure and function of a pentapeptide isolated from the gut of the eel.  
A:Reference number: JH0253; MUID:92062113  
A:Accession: JH0253  
A:Molecule type: protein  
A:Residues: 1-5 <UES>  
A:Experimental source: gut  
C:Comment: This peptide increased basal tone of the circular muscle of the esophagogastric, and of the circular muscle of the gastro-intestinal junction.

Query Match 30.4%; Score 14; DB 2; Length 5;  
Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 YW 7  
1 1  
Db 2 FW 3

## RESULT 10

B35640  
cerebellar degeneration-related protein - mouse (fragment)  
C:Species: *Mus musculus* (house mouse)  
C:Date: 28-Sep-1990 #sequence\_revision 28-Sep-1990 #text\_change 24-Jun-1993  
C:Accession: B35640  
R:Chen, Y.T.; Rettig, W.J.; Yenamandra, A.K.; Kozak, C.A.; Chaganti, R.S.K.; Posner, J.H.  
Proc. Natl. Acad. Sci. U.S.A. 87: 3077-3081, 1990  
A:Title: Cerebellar degeneration-related antigen: a highly conserved neuroectodermal marker  
A:Reference number: A35640; MUID:90222173  
A:Accession: B35640  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-6 <CHE>

Query Match 30.4%; Score 14; DB 2; Length 6;  
Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 YW 7  
1 1  
Db 1 FW 2

## RESULT 11

S21230  
dermorphin (Trp-4, Asn-7) [validated] - two-colored leaf frog (fragment)  
C:Species: *Phyllomedusa bicolor* (two-colored leaf frog)  
C:Date: 19-Mar-1997 #sequence\_revision 10-Oct-1997 #text\_change 18-Aug-2000  
C:Accession: S21230  
R:Mignogna, G.; Severini, C.; Simmaco, M.; Negri, L.; Falconieri Espamer, G.; Kreil, G.  
FEBS Lett. 302: 151-154, 1992  
A:Title: Identification and characterization of two dermorphins from skin extracts of the  
A:Reference number: S21230; MUID:92339502  
A:Accession: S21230  
A:Status: Preliminary  
A:Molecule type: protein

A:Residues: 1-7 <MG>  
C:Superfamily: dermorphin precursor; dermorphin precursor amino-terminal homology

Query Match 30.4%; Score 14; DB 2; Length 7;  
Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 YW 7  
1 1  
Db 3 FW 4

## RESULT 12

S33567  
tubulin beta-3 chain - fruit fly (*Drosophila melanogaster*) (fragment)  
C:Species: *Drosophila melanogaster*  
C:Date: 22-Nov-1993 #sequence\_revision 26-May-1995 #text\_change 21-Jul-2000  
C:Accession: S33567  
R:Rinz, U.; Wolk, A.; Renkawitz-Pohl, R.  
Development 116: 543-554, 1992  
A:Title: Ultrabithorax is a regulator of beta-3 tubulin expression in the *Drosophila*  
A:Reference number: S33567; MUID:93170162  
A:Accession: S33567  
A:Status: Preliminary; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-7 <HIN>  
A:Cross-references: EMBL:X68393; NID:97669; PIDN:CA448459.1; PID:94377448  
C:Genetics:  
A:Gene: *FlyBase:cbgr*  
A:Cross-references: *FlyBase:FBgn0003888*  
A:Introns: 5/3

Query Match 30.4%; Score 14; DB 2; Length 7;  
Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 YW 7  
1 1  
Db 6 FW 7

## RESULT 13

S22428  
chitin-binding protein - potato (fragment)  
C:Species: *Solanum tuberosum* (potato)  
C:Date: 19-Mar-1997 #sequence\_revision 10-Oct-1997 #text\_change 05-Dec-1998  
C:Accession: S22428  
R:Millar, D.J.; Allen, A.K.; Smith, C.G.; Sidebottom, C.; Slabas, A.R.; Botwell, G.P.  
Biochem. J. 283: 813-821, 1992  
A:Title: Chitin-binding proteins in potato (*Solanum tuberosum* L.) tuber. Characteriza  
A:Reference number: S21288; MUID:92272683  
A:Accession: S22428  
A:Molecule type: protein  
A:Residues: 1-8 <ML>  
C:Function:  
A:Description: may be involved in plant defence  
C:Keywords: glycoprotein; hydroxyproline

Query Match 30.4%; Score 14; DB 2; Length 8;  
Best Local Similarity 33.3%; Pred. No. 2.8e+05;  
Matches 2; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 PKLTY 6  
1 1  
Db 1 PKDMY 6

## RESULT 14

T49424  
cytotoxic T-lymphocyte proteinase 3 (EC 3.4.21.-) - western wild mouse (fragment)

C:Species: Mus spretus (western wild mouse)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
C:Accession: 149424  
R:Ko, M.S.; Wang, X.; Horton, J.H.; Hagen, M.D.; Takahashi, N.; Maezaki, Y.; Nadeau, J.H.  
Mamm. Genome 5, 349-355, 1994  
A:Title: Genetic mapping of 40 cDNA clones on the mouse genome by PCR.  
A:Reference number: I48934; MUID:94319082  
A:Accession: 149424  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-6 <RES>  
A:Cross-references: EMBL:U05745; NID:9497084; PIDN:AAB60461.1; PID:9642831  
C:Keywords: hydrolase; serine proteinase

Query Match 28.3%; Score 13; DB 2; Length 6;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KLL 4  
|||  
Db 4 KLL 6

RESULT 15  
S33246  
neuromodulatory peptide Wwamide-3 - giant African snail  
C:Species: Achatina fulica (giant African snail)  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 24-Jul-1997  
C:Accession: S33246  
R:Minakata, H.; Ikeda, T.; Muneoka, Y.; Kobayashi, M.; Nomoto, K.  
FEBS Lett. 323, 104-108, 1993  
A:Title: Wwamide-1, -2 and -3: novel neuromodulatory peptides isolated from ganglia of t  
A:Reference number: S33244; MUID:93265812  
A:Accession: S33246  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-7 <MIN>

Query Match 28.3%; Score 13; DB 2; Length 7;  
Best Local Similarity 33.3%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 KLLTY 7  
| : |  
Db 2 KQMSV 7

Search completed: July 15, 2002, 13:39:06  
Job time: 236 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 15, 2002, 13:39:10 ; Search time 10.16 Seconds  
(without alignments)  
30.488 Million cell updates/sec

Title: US-09-712-819A-4

Sequence: 46  
1 PKLLIYMA 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 148

Minimum DB seq length: 0  
Maximum DB seq length: 8

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	37.0	7	1	GPRP_MOUSE
2	15	32.6	5	1	BPP7_BOTIN
3	13	28.3	7	1	WMA2_ACHFU
4	13	28.3	8	1	AC1_TFUAL
5	13	28.3	8	1	LMT2_LOCM1
6	13	28.3	8	1	LPR_LEUMA
7	13	28.3	8	1	PPK2_PERAM
8	13	28.3	8	1	PPK3_PERAM
9	12	26.1	7	1	TVS1_LITRU
10	12	26.1	7	1	UP04_MOUSE
11	12	26.1	7	1	WMA3_ACHFU
12	12	26.1	8	1	RS7_MYCIT
13	11	23.9	4	1	OCB3_OCTMI
14	11	23.9	4	1	PSK_DAUCA
15	11	23.9	5	1	UF01_MOUSE
16	11	23.9	6	1	EI01_LITRU
17	11	23.9	6	1	LOK1_LOCM1
18	11	23.9	7	1	EI05_LITRU
19	11	23.9	7	1	HY7_PIG
20	11	23.9	7	1	WMA1_ACHFU
21	11	23.9	8	1	AKHG_GRYBI
22	11	23.9	8	1	AKH_LITRU
23	11	23.9	8	1	AKH_MEML
24	11	23.9	8	1	AKH_TYBAT
25	11	23.9	8	1	CKKN_MACED
26	11	23.9	8	1	FAR1_PANRE
27	11	23.9	8	1	FAR1_ASCSU
28	11	23.9	8	1	HTEP2_PERAM
29	11	23.9	8	1	HTEP2_PERAM
30	11	23.9	8	1	HTEP2_PERAM
31	11	23.9	8	1	HTEP2_PERAM
32	11	23.9	8	1	LCK1_LEUMA
33	11	23.9	8	1	LCK3_LEUMA
34	11	23.9	8	1	LCK3_LEUMA

34	11	23.9	8	1	LCK4_LEUMA	P21143 leucophaea
35	11	23.9	8	1	LCK5_LEUMA	P19987 leucophaea
36	11	23.9	8	1	LCK6_LEUMA	P19988 leucophaea
37	11	23.9	8	1	LCK7_LEUMA	P19989 leucophaea
38	11	23.9	8	1	LCK8_LEUMA	P19990 leucophaea
39	11	23.9	8	1	PLP_BRANA	P11707 brassica na
40	11	23.9	8	1	RECH_PAMBO	P08939 pandanus bo
41	10	21.7	7	1	CARP_MYTED	P10420 mytilus edu
42	10	21.7	7	1	CCCF1_ENTFA	P20104 enterococcu
43	10	21.7	8	1	ALL6_CYPDO	P20157 cydia pomon
44	10	21.7	8	1	ANG2_BORJA	Q10582 bothrops ja
45	10	21.7	8	1	UPAA_HUMAN	P30096 homo sapien

## ALIGNMENTS

RESULT 1  
GPRP\_MOUSE  
ID GPRP\_MOUSE STANDARD; PRT; 7 AA.  
AC P99025;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE GPR cyclohydrolase I feedback regulatory protein (P35) (Fragment).  
GN GCHFR OR GPRP.  
OS Mus musculus (mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID:10090;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Liver;  
RA Sanchez J.-C., Rouge V., Frutiger S., Hughes G., Yan J.X.,  
RA Hoogland C., Appel R.D., Binz P.A., Hochstrasser D.F.,  
RA Cowthorne M.;  
RL Submitted (Aug-1998) to the SWISS-PROT data bank.  
CC -!- FUNCTION: MEDIATES TETRAHYDROBIOTERIN INHIBITION OF GPR  
CYCLOHYDROLASE I. THIS INHIBITION IS REVERSED BY L-PHENYLMALANINE  
(BY SIMILARITY).  
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC SWISS-2DPAGE: P99025; MOUSE.  
DR INIT\_MET 0  
FT NON\_TER 7  
SQ SEQUENCE 7 AA: 806 MW: 718505727384700 CRC64;

Query Match 37.0%; Score 17; DB 1; Length 7;  
Best Local Similarity 80.0%; Pred. No. 1e+05;  
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKLLI 5  
1 1 1 1 1  
DB 1 PYLLI 5

RESULT 2  
BPP7\_BOTIN  
ID BPP7\_BOTIN STANDARD; PRT; 5 AA.  
AC P30425;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE Bradykinin-potentiating peptide S5.2 (5a) (Angiotensin-converting  
enzyme inhibitor).  
OS Eukaryota; Metazoa; (Island Jarraca) (Queimada Jarraca).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Lepidodactylus; Squamata; Scleroglossa; Serpentes; Colubroidae;  
OC Viperidae; Crotalinae; Bothrops.  
OX NCBI\_TaxID=8723;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Venom;

RX MEDLINE-9035157; PubMed-2386615;  
 RA Cindra A.C.O., Vieira C.A., Giglio J.R.;  
 RT "Primary structure and biological activity of bradykinin potentiating  
 RT peptides from Bothrops insularis snake venom."  
 RL J. Protein Chem. 9:221-227(1990).  
 CC -1- FUNCTION: THIS PEPTIDE BOTH INHIBITS THE ACTIVITY OF THE  
 CC ANGIOTENSIN-CONVERTING ENZYME AND ENHANCES THE ACTION OF  
 CC BRADYKININ BY INHIBITING THE KINASES THAT INACTIVATE IT.  
 CC IT ACTS AS AN INDIRECT HYPOTENSIVE AGENT.  
 DR PIR: G37196; G37196.  
 KM Hypotensive agent; Venom.  
 FT MOD.RES 1  
 SQ SEQUENCE 5 AA: 629 MW: 776DC37326B0000 CRC64.

Query Match 32.6%; Score 15; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 WA 8  
 DB 3 WA 4

RESULT 3  
 WMA2\_ACHFU STANDARD; PRT; 7 AA.  
 ID WMA2\_ACHFU  
 AC P33920;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last annotation update)  
 DE Wamide-2.  
 OS Achatina fulica (Giant African snail).  
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;  
 OC Achatinacea; Achatinidae; Achatina.  
 OX NCBI\_TaxID=6530;  
 RN (1)  
 RP SEQUENCE.  
 RC TISSUE=ganglion;  
 RX MEDLINE-93265912; PubMed-8495720;  
 RA Minakata H., Ikeda T., Muneeoka Y., Kobayashi M., Nomoto K.;  
 RT "Wamide-1, -2 and -3: novel neuromodulatory peptides isolated from  
 RT ganglia of the African giant snail, Achatina fulica."  
 RL FEBS Lett. 323:104-108(1993).  
 KM Neuropeptide; Amidation.  
 FT MOD.RES 7  
 SQ SEQUENCE 7 AA: 964 MW: 7362D5B686D32310 CRC64;

Query Match 28.3%; Score 13; DB 1; Length 7;  
 Best Local Similarity 33.3%; Pred. No. 1e+05;  
 Matches 2; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 KLLIYW 7  
 DB 2 KQMSYV 7

RESULT 4  
 ACT\_THUAL STANDARD; PRT; 8 AA.  
 ID ACT\_THUAL  
 AC P18691;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-NOV-1990 (Rel. 16, Last annotation update)  
 DE Angiotensin-converting enzyme inhibitor.  
 OS Thunus albacares (Yellowfin tuna) (Neohunus macropterus).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostei;  
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 CC Acanthomorpha; Acanthopterygii; Perciformes; Scombroidei;  
 OC Scombridae; Thunnus.  
 OX NCBI\_TaxID=8236;  
 RN (1)

RP SEQUENCE.  
 RC TISSUE=Muscle;  
 RX MEDLINE-88326322; PubMed-3415688;  
 RA Kohama Y., Matsumoto S., Oka H., Teramoto T., Okabe M., Minura T.;  
 RT "Isolation of angiotensin-converting enzyme inhibitor from tuna  
 RT muscle."  
 RL Biochem. Biophys. Res. Commun. 155:332-337(1988).  
 DR PIR: A31570; A31570.  
 SQ SEQUENCE 8 AA: 953 MW: 6AA863733051F1B7 CRC64;

Query Match 28.3%; Score 13; DB 1; Length 8;  
 Best Local Similarity 66.7%; Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 IYW 7  
 DB 4 IKW 6

RESULT 5  
 LMT2\_LOCMI STANDARD; PRT; 8 AA.  
 ID LMT2\_LOCMI  
 AC P22396;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 01-AUG-1991 (Rel. 19, Last annotation update)  
 DE Locustamylotropin 2 (LWM-MT-2).  
 OS Locusta migratoria (Migratory locust).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caellifera;  
 OC Acridoidea; Acrididae; Locusta.  
 OX NCBI\_TaxID=7004;  
 RN (1)  
 RP SEQUENCE.  
 RC TISSUE=Corpora cardiaca;  
 RA Schoofs L., Holman G.M., Hayes T.K., Nachman R.J., de loof A.;  
 RT "Isolation, identification and synthesis of locustamylotropin II, an  
 RT additional neuropeptide of Locusta migratoria. Member of the  
 RT cephalomylotropic peptide family."  
 RL Insect Biochem. 20:479-484(1990).  
 CC -1- FUNCTION: MEDIATES VISCERAL MUSCLE CONTRACTILE ACTIVITY  
 CC (MYOTROPIC ACTIVITY).  
 CC -1- SIMILARITY: BELONGS TO THE PYROKININ FAMILY.  
 CC Interpro: IPR001484; Pyrokinin.  
 DR Prosite: PS00539; Pyrokinin; 1.  
 KM Neuropeptide; Amidation; Pyrokinin.  
 FT MOD.RES 8  
 SQ SEQUENCE 8 AA: 934 MW: 26341771A9CA87B CRC64;

Query Match 28.3%; Score 13; DB 1; Length 8;  
 Best Local Similarity 66.7%; Pred. No. 1e+05;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKL 3  
 DB 6 PRL 8

RESULT 6  
 LPK\_LEUMA STANDARD; PRT; 8 AA.  
 ID LPK\_LEUMA  
 AC P13049;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Leukopyrokinin (LPK) (LEW-PK).  
 OS Leucophaea maderae (Madeira cockroach).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Orthopteroidea; Dicoptera; Blattaria;  
 OC Blaberidae; Blaberidae; Leucophaea.  
 OX NCBI\_TaxID=6988;

RN SEQUENCE [1]  
 RP MEDLINE=86269041; PubMed=3015140;  
 RA Nachman R.J., Holman G.M., Cook B.J.;  
 RT "Active fragments and analogs of the insect neuropeptide  
 leucopyrokinin: structure-function studies.";  
 RL Biochem. Biophys. Res. Commun. 137:936-942(1986).  
 RN [2]  
 RP SEQUENCE, AND SYNTHESIS.  
 RC TISSUE-Head;  
 RX MEDLINE=87052651; PubMed=2877794;  
 RA Holman G.M., Cook B.J., Nachman R.J.;  
 RT "Primary structure and synthesis of a blocked myotropic  
 neuropeptide isolated from the cockroach, *Leucophaea maderae*.";  
 RL Comp. Biochem. Physiol. 85C:219-224(1986).  
 CC -1- FUNCTION: MEDIATES VISCERAL MUSCLE CONTRACTILE ACTIVITY  
 (MYOTROPIC ACTIVITY).  
 CC -1- MISCELLANEOUS: AN ANALOG WITHOUT THE N-TERMINAL PCA RESIDUE WAS  
 SYNTHESIZED AND FOUND TO EXHIBIT GREATER ACTIVITY (144%) THAN THE  
 PARENT NEUROPEPTIDE. THE PORTION OF THE SEQUENCE OF LFK MOST  
 CRITICAL FOR THE MYOTROPIC PROPERTIES IS LIMITED TO THE  
 PENTAPEPTIDE FRAGMENT FPPRL.  
 CC -1- SIMILARITY: BELONGS TO THE PYROKININ FAMILY.  
 DR InterPro: IPR001484; Pyrokinin.  
 DR PROSITE: PS00539; PYROKININ: 1.  
 KM Neuropeptide: Amidation; Pyrokinin.  
 FT MOD\_RES 1 1 PYROKINONE CARBOXYLIC ACID.  
 FT MOD\_RES 8 8 AMIDATION.  
 SQ SEQUENCE 8 AA: 949 MW; 9234171A9D5A1B6 CRC64;

Query Match 28.3%; Score 13; DB 1; Length 8;  
 Best Local Similarity 66.7%; Pred. No. 1e+05;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKL 3  
 1:1  
 DB 6 PRL 8

RESULT 7  
 ID PPK3\_PERAM STANDARD; PRT; 8 AA.  
 AC P82618;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Pyrokinin-3 (Pep-PK-3) (FPPRL-amide)  
 OS *Periplaneta americana* (American cockroach).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;  
 CC Blattodea; Blattidae; Periplaneta.  
 OC NCBI\_TaxID=6978;  
 RN [1]  
 RP SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.  
 RC TISSUE-Corpus cardiaca;  
 RX MEDLINE=97353923; PubMed=9210163;  
 RA Predel R., Kellner R., Kaufmann R., Penzlin H., Gaede G.;  
 RT "Isolation and structural elucidation of two pyrokinins from the  
 retrocerebral complex of the American cockroach.";  
 RL Peptides 18:473-478(1997).  
 RN [2]  
 RP TISSUE SPECIFICITY.  
 RX MEDLINE=20189894; PubMed=10723010;  
 RA Predel R., Eckert M.;  
 RT "Tegma-specific distribution of Fxprlamides in the nervous system of  
 the American cockroach.";  
 RL J. Comp. Neurol. 419:352-363(2000).  
 CC -1- FUNCTION: MEDIATES VISCERAL MUSCLE CONTRACTILE ACTIVITY  
 (MYOTROPIC ACTIVITY).  
 CC -1- TISSUE SPECIFICITY: CORPORA CARDIACA.  
 CC -1- MASS SPECTROMETRY: MM=883; METHOD=MALDI.

CC -1- SIMILARITY: BELONGS TO THE PYROKININ FAMILY.  
 DR InterPro: IPR001484; Pyrokinin.  
 DR PROSITE: PS00539; PYROKININ: FALSE\_NEG.  
 KM Neuropeptide: Amidation; Pyrokinin.  
 FT MOD\_RES 8 8 AMIDATION.  
 SQ SEQUENCE 8 AA: 884 MW; C834176DD9D7775 CRC64;

Query Match 28.3%; Score 13; DB 1; Length 8;  
 Best Local Similarity 66.7%; Pred. No. 1e+05;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKL 3  
 1:1  
 DB 6 PRL 8

RESULT 8  
 ID PPK3\_PERAM STANDARD; PRT; 8 AA.  
 AC P82618;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Pyrokinin-3 (Pep-PK-3) (FPPRL-amide).  
 OS *Periplaneta americana* (American cockroach).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;  
 CC Blattodea; Blattidae; Periplaneta.  
 OC NCBI\_TaxID=6978;  
 RN [1]  
 RP SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.  
 RC TISSUE-Retrocerebral complex;  
 RX MEDLINE=99212469; PubMed=10196736;  
 RA Predel R., Kellner R., Nachman R.J., Holman G.M., Rapus J., Gaede G.;  
 RT "Differential distribution of pyrokinin-isoforms in cerebral and  
 abdominal neurohemal organs of the American cockroach.";  
 RL Insect Biochem. Mol. Biol. 29:139-144(1999).  
 RN [2]  
 RP TISSUE SPECIFICITY.  
 RX MEDLINE=20189894; PubMed=10723010;  
 RA Predel R., Eckert M.;  
 RT "Tegma-specific distribution of Fxprlamides in the nervous system of  
 the American cockroach.";  
 RL J. Comp. Neurol. 419:352-363(2000).  
 CC -1- FUNCTION: MEDIATES VISCERAL MUSCLE CONTRACTILE ACTIVITY  
 (MYOTROPIC ACTIVITY).  
 CC -1- TISSUE SPECIFICITY: CORPORA CARDIACA.  
 CC -1- MASS SPECTROMETRY: MM=996.5; METHOD=MALDI.  
 CC -1- SIMILARITY: BELONGS TO THE PYROKININ FAMILY.  
 KM Neuropeptide: Amidation; Pyrokinin.  
 FT MOD\_RES 8 8 AMIDATION.  
 SQ SEQUENCE 8 AA: 997 MW; 0B3417409D772C7 CRC64;

Query Match 28.3%; Score 13; DB 1; Length 8;  
 Best Local Similarity 66.7%; Pred. No. 1e+05;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKL 3  
 1:1  
 DB 6 PRL 8

RESULT 9  
 ID TYS1\_LITRU STANDARD; PRT; 7 AA.  
 AC P82065;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DE Trypophyllin 5.1.  
 OS *Litoria rubella* (Desert tree frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
 OC Litoria;  
 OX NCBI\_TaxID=104895;  
 RN [1]  
 RP SEQUENCE, AND MASS SPECTROMETRY.  
 RC TISSUE-SKIN SECRETION;  
 RA Steinhorn S.T., Wehlitz P.A., Maugh R.J., Bowle J.H., Gao C.,  
 RA Tyler M.J., Wallace J.C.;  
 RT 'The structure of new peptides from the Australian red tree frog  
 RT 'Litoria rubella'. The skin peptide profile as a probe for the study  
 RT of evolutionary trends of amphibians.';  
 RL Anst. J. Chem. 49:955-963(1996).  
 CC -1- FUNCTION: MAY ACT AS A NEUROMODULATOR OR NEUOTRASMITTER.  
 CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
 CC Amphibian skin; Amidation: Neuropeptide.  
 KW MOD\_RES 1 7  
 FT MOD\_RES 7 7  
 FT NON\_TER 7 7  
 SQ SEQUENCE 7 AA: 983 MW: 74019D3676046B0 CRC64:

Query Match 26.1%; Score 12; DB 1; Length 7;  
 Best Local Similarity 66.7%; Pred. NO. 1e+05;  
 Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 IYW 7  
 Db 2 IPW 4

RESULT 10  
 ID UP04\_MOUSE STANDARD; PRT: 7 AA.  
 AC P38642;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)  
 DE Unknown protein from 2D-page of fibroblasts (P46) (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE-FIBROBLAST;  
 RX MEDLINE=95009907; PubMed=7523108;  
 RA Merrick B.A., Patterson R.M., Wichter L.L., He C., Selkirk J.K.;  
 RT "Separation and sequencing of familial and novel murine proteins  
 RT using preparative two-dimensional gel electrophoresis.";  
 RL Electrophoresis 15:735-745(1994).  
 CC -1- MISCELLANEOUS: ON THE 2D-GEL, THE DETERMINED PI OF THIS UNKNOWN  
 CC PROTEIN IS: 5.0, ITS MW IS: 46 kDa.  
 FT NON\_TER 7 7  
 SQ SEQUENCE 7 AA: 766 MW: 68640AB77632700 CRC64:

Query Match 26.1%; Score 12; DB 1; Length 7;  
 Best Local Similarity 100.0%; Pred. NO. 1e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PK 2  
 Db 1 PK 2

RESULT 11  
 ID WMA3\_ACHFU STANDARD; PRT: 7 AA.  
 AC P35921;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last annotation update)

DE Wamide-3.  
 OS Achatina fulica (Giant African snail).  
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;  
 OC Achatinacea; Achatinidae; Achatina.  
 OX NCBI\_TaxID=6530;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE-GANGLION;  
 RX MEDLINE=93265912; PubMed=8495720;  
 RA Minakata H., Ikeda T., Munooka Y., Kobayashi M., Nomoto K.;  
 RT "Wamide-1, -2 and -3: novel neuromodulatory peptides isolated from  
 RT ganglia of the African giant snail, Achatina fulica.";  
 RL FEBS Lett. 323:104-108(1993).  
 DR PIR: S33244; S33244.  
 KW Neuropptide; Amidation.  
 FT MOD\_RES 7 7  
 FT NON\_TER 7 7  
 SQ SEQUENCE 7 AA: 965 MW: 7362D5B69B132310 CRC64:

Query Match 26.1%; Score 12; DB 1; Length 7;  
 Best Local Similarity 33.3%; Pred. NO. 1e+05;  
 Matches 2; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 KLIYW 7  
 Db 2 KEMSW 7

RESULT 12  
 ID RS7\_MYCIT STANDARD; PRT: 8 AA.  
 AC P33564;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE 30S ribosomal protein S7 (Fragment).  
 GN RPSG.  
 OS Mycobacterium intracellulare.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1767;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93197130; PubMed=8451173;  
 RA Nair J., Rouse D.A., Morris S.L.;  
 RT "Nucleotide sequence analysis of the ribosomal S12 gene of  
 RT Mycobacterium intracellulare.";  
 RL Nucleic Acids Res. 21:1038-1039(1993).  
 CC -1- FUNCTION: PROTEIN S7 BINDS SPECIFICALLY TO PART OF THE 3' END OF  
 CC 16S RIBOSOMAL RNA (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE S7P FAMILY OF RIBOSOMAL PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 CC -----  
 CC EMBL: L08171; AAA25376.1; -.  
 DR PIR: S35538; S35538.  
 DR InterPro: IPR000235; Ribosomal\_S7.  
 DR PROSITE: PS00052; RIBOSOMAL\_S7; PARTIAL.  
 KW Ribosomal protein; RNA-binding.  
 FT INIT\_MET 0 0  
 FT NON\_TER 8 8  
 SQ SEQUENCE 8 AA: 850 MW: 63276DC768732417 CRC64:

Query Match 26.1%; Score 12; DB 1; Length 8;  
 Best Local Similarity 100.0%; Pred. NO. 1e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PK 2  
11  
Db 7 PK 8

RESULT 13  
OCP3\_OCTMI STANDARD: PRT: 4 AA.

AC P58649;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Cardioactive peptides Ocp-3/Ocp-4.  
OS Octopus minor (Octopus).  
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Octopoda;  
OC Invertebrata; Octopodidae; Octopus.  
OX NCBI\_TaxID=89766;  
RN [1]  
RP SEQUENCE, SYNTHESIS, MASS SPECTROMETRY, AND CHARACTERIZATION.

RC TISSUE-Brain;  
RX PubMed=10876044;  
RA Iwakoshi E., Hisada M., Minakata H.;  
RT "Cardioactive peptides isolated from the brain of a Japanese octopus,  
Octopus minor".  
RL Peptides 21:623-630(2000).  
CC -1- FUNCTION: Cardioactive; has both positive chronotropic and  
inotropic effects on the heart. Ocp-4 is a 1000 time less  
active than Ocp-3.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- PTM: Ocp-4 has D-Ser instead of L-Ser.  
CC -1- MASS SPECTROMETRY: MW=395.2; METHOD=MALDI.  
KW Hormone; D-amino acid.  
FT MOD\_RES 2 D-SERINE (IN OCP-4).  
SQ SEQUENCE 4 AA; 463 MW; 6AB365B810000000 CRC64;

Query Match 23.9%; Score 11; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 W 7  
1  
Db 3 W 3

RESULT 14  
PSK\_DAUCA STANDARD: PRT: 5 AA.

AC P58261;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Phytosulfokine-alpha (PSK-alpha) [Contains: Phytosulfokine-beta (PSK-beta)].  
OS Daucus carota (Carrot).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; easterids II; Apiales; Apiaceae; Daucus.  
OX NCBI\_TaxID=4039;  
RN [1]  
RP SEQUENCE, AND IDENTIFICATION BY MASS SPECTROMETRY.

RC STRAIN=cv US-Harumakigusun;  
RX MEDLINE=20212743; PubMed=10750705;  
RA Hanai H., Matsuno T., Yamamoto M., Matsubayashi Y., Kobayashi T.,  
Kamada H., Sakagami Y.;  
RT "A secreted peptide growth factor, phytosulfokine, acting as a  
stimulatory factor of carrot somatic embryo formation.";  
RL Plant Cell Physiol. 41:27-32(2000).  
CC -1- FUNCTION: IN PRESENCE OF 2,4-D, STIMULATES PROLIFERATION OF THE  
CELLS. BUT DOES NOT STIMULATE DIFFERENTIATION INTO THE SOMATIC  
EMBRYOS  
CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- PTM: SULFATION IS IMPORTANT FOR ACTIVITY AND FOR THE BINDING TO A  
PUTATIVE MEMBRANE RECEPTOR (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE PHYTOSULFOKINE FAMILY.  
KW Growth factor; Sulfation.  
FT PEPTIDE 1 4 PHYTOSULFOKINE-BETA.  
FT MOD\_RES 1 1 SULFATION.  
FT MOD\_RES 3 3 SULFATION.  
SQ SEQUENCE 5 AA; 687 MW; 76C1BB504B300000 CRC64;

Query Match 23.9%; Score 11; DB 1; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 IV 6  
11  
Db 2 IV 3

RESULT 15  
UF01\_MOUSE STANDARD: PRT: 5 AA.

AC P38639;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE Unknown protein from 2D-page of fibroblasts (P19) (Fragment).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Fibroblast;  
RX MEDLINE=95009907; PubMed=7523108;  
RA Merrick B.A., Patterson R.M., Wilcher L.L., He C., Selkirk J.K.;  
RT "Separation and sequencing of familial and novel murine proteins  
using preparative two-dimensional gel electrophoresis.";  
RL Electrophoresis 15:735-745(1994).  
CC -1- MISCELLANEOUS: ON THE 2D-GEL, THE DETERMINED PI OF THIS UNKNOWN  
PROTEIN IS: 6.6, ITS MW IS: 19 KDa.  
FT NON\_TER 5  
SQ SEQUENCE 5 AA; 717 MW; 7364087043100000 CRC64;

Query Match 23.9%; Score 11; DB 1; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1e+05;  
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 W 7  
1  
Db 1 W 1

Search completed: July 15, 2002, 13:42:21  
Job time: 191 sec

10



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:38:50 ; Search time 23.77 Seconds

(Without alignments)  
58.223 Million cell updates/sec

Title: US-09-712-819A-4

Sequence: 1 PKLITYMA 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 378

Minimum DB seq length: 0

Maximum DB seq length: 8

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

SPPREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_podent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rv1rus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	37.0	8	Q9R049	Q9R049 Buchnera ap
2	16	34.8	8	O85406	O85406 coxiella bu
3	15	32.6	8	O94VJ4	O94VJ4 varanus ben
4	15	32.6	8	O94V82	O94V82 varanus yuv
5	14	30.4	8	O42507	O42507 tritlicum ae
6	13	28.3	7	O50556	O50556 actinobacil
7	13	28.3	7	O66205	O66205 transmissib
8	13	28.3	8	O87471	O87471 haemophilus
9	13	28.3	8	O02831	O02831 oryctolagus
10	12	26.1	7	P92421	P92421 psathyrosta
11	12	26.1	7	P92385	P92385 hordeum mar
12	12	26.1	7	P92372	P92372 haynaldia v
13	12	26.1	7	P92403	P92403 lophopyrum
14	12	26.1	7	P92425	P92425 pseudorogn
15	12	26.1	7	P92387	P92387 hennardia p
16	12	26.1	8	P92427	P92427 peridictyon

17	12	26.1	7	P92390	P92390 heteranthe
18	12	26.1	7	P92226	P92226 crithopsis
19	12	26.1	7	P92214	P92214 amblyopyrum
20	12	26.1	7	P92430	P92430 begliops ta
21	12	26.1	7	P92221	P92221 bromus iner
22	12	26.1	7	P92442	P92442 taenialtheru
23	12	26.1	7	P92381	P92381 hordeum bra
24	12	26.1	7	P92393	P92393 hordeum vul
25	12	26.1	7	P92218	P92218 australopyr
26	12	26.1	7	P92440	P92440 thnopyrum
27	12	26.1	7	P92210	P92210 agropyron c
28	12	26.1	7	P82065	P82065 ilitoria rub
29	12	26.1	8	O921E9	O921E9 nelsseria m
30	12	26.1	8	O9R057	O9R057 buchnera ap
31	12	26.1	8	O15888	O15888 homo sapien
32	12	26.1	8	O19956	O19956 gossypium d
33	12	26.1	8	O19957	O19957 gossypium a
34	12	26.1	8	O19958	O19958 gossypium b
35	12	26.1	8	O19959	O19959 gossypium t
36	12	26.1	8	O19960	O19960 gossypium m
37	12	26.1	8	O19961	O19961 gossypium d
38	12	26.1	8	O94VC1	O94VC1 varanus rud
39	12	26.1	8	O94VB5	O94VB5 varanus sal
40	12	26.1	8	O94VB2	O94VB2 varanus sal
41	12	26.1	8	O94VA7	O94VA7 varanus sal
42	12	26.1	8	O94PX7	O94PX7 felis silve
43	12	26.1	8	O94PX6	O94PX6 felis libyc
44	12	26.1	8	O94PX5	O94PX5 felis silve
45	12	26.1	11	P82598	P82598 rattus norv

#### ALIGNMENTS

RESULT 1  
ID Q9R049 PRELIMINARY: PRT: 8 AA.  
AC Q9R049  
DT 01-MAY-2000 (TREMREL. 13, Created)  
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)  
DT 01-MAY-2000 (TREMREL. 13, Last annotation update)  
DE NIFS PROTEIN HOMOLOG (FRAGMENT).  
GN NIFS.  
OS Buchnera aphidicola.  
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.  
OX NCBI\_Taxid-9;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-20022990; PubMed-10555290;  
RA Clark M.A., Moran N.A., Baumann P.;  
RT "Sequence evolution in bacterial endosymbionts having extreme base  
composition.";  
RT MOI. Biol. Evol. 16:1586-1598(1999).  
DR EMBL, AF130814; AAF13805.1; .  
FT NON\_TER  
SQ SEQUENCE 8 AA; 992 MW; F3A73B5047672336 CRC64;

Query Match 37.0%; Score 17; DB 2; Length 8;  
Best Local Similarity 80.0%; Pred. No. 5.6e+05;  
Matches 4; Conservative 0; Mismatches 1; Indels 0;  
Gaps 0;  
QY 2 KLIV 6  
DB 2 KLIV 6  
RESULT 2  
ID O85406 PRELIMINARY: PRT: 8 AA.  
AC O85406  
DT 01-NOV-1998 (TREMREL. 08, Created)  
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 1.0 KDA PROTEIN (FRAGMENT).  
 OS Coxiella burnetii.  
 OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;  
 OC Coxiella group; Coxiella.  
 OX NCBI\_TaxID=777;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NINE MILE PHASE I;  
 RX MEDLINE=98348442; PubMed=9683477;  
 RA Williams H., Jeeger C., Baljer G.;  
 RT "Physical and genetic map of the obligate intracellular bacterium  
 RT Coxiella burnetii".  
 RL J Bacteriol 180:3816-3822(1998).  
 DR EMBL AF064963; AAD0947.1; -.  
 KW Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 8 AA; 993 MW; 046B5AA453772727 CRC64;

Query Match 34.8%; Score 16; DB 2; Length 8;  
 Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LTYW 7  
 DB 1 LLMW 4

RESULT 3  
 ID 094VJ4 PRELIMINARY; PRT; 8 AA.  
 AC 094VJ4;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CYTOCHROME C OXIDASE SUBUNIT I (FRAGMENT).  
 GN COI.  
 OS Varanus bengalensis nebulosis.  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lepidosaurs; Squamata; Scleroglossa; Anguimorpha; Varanidae; Varanus.  
 OX NCBI\_TaxID=169827;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ast J.C.;  
 RT "Mitochondrial DNA evidence and evolution in Varanoidea (Squamata)".  
 RL Cladistics 17:0-0(2001).  
 DR EMBL AF407492; AAL10031.1; -.  
 KW Mitochondrion.  
 FT NON\_TER  
 SQ SEQUENCE 8 AA; 1053 MW; E8B5B9C733640056 CRC64;

Query Match 32.6%; Score 15; DB 8; Length 8;  
 Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LTYW 7  
 DB 1 MIRM 4

RESULT 4  
 ID 094V82 PRELIMINARY; PRT; 8 AA.  
 AC 094V82;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CYTOCHROME C OXIDASE SUBUNIT I (FRAGMENT).  
 GN COI.  
 OS Varanus yunnanol.

OC Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lepidosaurs; Squamata; Scleroglossa; Anguimorpha; Varanidae; Varanus.  
 OX NCBI\_TaxID=169856;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ast J.C.;  
 RT "Mitochondrial DNA evidence and evolution in Varanoidea (Squamata)".  
 RL Cladistics 17:0-0(2001).  
 DR EMBL AF407535; AAL10157.1; -.  
 KW Mitochondrion.  
 FT NON\_TER  
 SQ SEQUENCE 8 AA; 1045 MW; EFC775A6C3640056 CRC64;

Query Match 32.6%; Score 15; DB 8; Length 8;  
 Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LTYW 7  
 DB 1 MIRM 4

RESULT 5  
 ID 042507 PRELIMINARY; PRT; 8 AA.  
 AC 042507;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE HEAT SHOCK PROTEIN (FRAGMENT).  
 GN HSP70C OR HSP70A OR HSP70B.  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticaceae; Triticum.  
 OX NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV M0STANG;  
 RX MEDLINE=96189275; PubMed=8605312;  
 RA Joshi C.P., Kumar S., Nguyen H.T.;  
 RT "Application of modified differential display technique for cloning  
 RT and sequencing of the 3' region from three putative members of wheat  
 RT HSP70 gene family".  
 RL Plant Mol. Biol. 30:641-646(1996).  
 DR EMBL L41507; AAB02333.1; -.  
 DR EMBL L41505; AAB02331.1; -.  
 DR EMBL L41506; AAB02332.1; -.  
 KW Heat shock.  
 FT NON\_TER  
 SQ SEQUENCE 8 AA; 886 MW; 71B2CB1B10532768 CRC64;

Query Match 30.4%; Score 14; DB 10; Length 8;  
 Best Local Similarity 66.7%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PXL 3  
 DB 2 PKI 4

RESULT 6  
 ID 050556 PRELIMINARY; PRT; 7 AA.  
 AC 050556;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLYA (FRAGMENT).  
 GN GLYA.

OS Actinobacillus actinomycetemcomitans (Haemophilus  
 OS actinomycetemcomitans).  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Actinobacillus.  
 OX NCBI\_TaxID=714;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 33384;  
 RX MEDLINE=96355846; PubMed=8751884;  
 RA Kolodrubetz D., Spltznagel J. Jr., Wang B., Phillips L.H., Jacobs C.,  
 RA Kraig E.;  
 RT "cis Elements and trans factors are both important in strain-specific  
 RT regulation of the leukotoxin gene in Actinobacillus  
 RT actinomycetemcomitans.";  
 RL Infect Immun. 64:3451-3460(1996).  
 DR EMBL: U51862; AAB8721.1; -;  
 FT NON\_TER  
 SQ SEQUENCE 7 AA; 832 MW; 6DCH42D767340420 CRC64;

Query Match 28.3%; Score 13; DB 2; Length 7;  
 Best Local Similarity 40.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 KILL 6  
 : 1 :  
 DB 2 RLPV 6

RESULT 7  
 ID 066205 PRELIMINARY; PRT; 7 AA.  
 AC 066205;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MEMBRANE PROTEIN (1 IS 3RD BASE IN CODON) (FRAGMENT).  
 OS Transmissible gastroenteritis virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronaviruses.  
 OX NCBI\_TaxID=11149;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FS772/70;  
 RX MEDLINE=88216185; PubMed=2835592;  
 RA Britton P., Carnes R.S., Page K.W., Garves D.J., Parra F.;  
 RT "Sequence of the Nucleoprotein Gene from a Virulent British Field  
 RT Isolate of Transmissible Gastroenteritis Virus and Its Expression in  
 RT Saccharomyces Cerevisiae.";  
 RL Mol. Microbiol. 2:89-99(1988).  
 DR EMBL: Y00542; CA68606.1; -;  
 FT NON\_TER  
 SQ SEQUENCE 7 AA; 869 MW; 72C691E727233B70 CRC64;

Query Match 28.3%; Score 13; DB 12; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KILL 4  
 : 1 :  
 DB 2 KILL 4

RESULT 8  
 ID 087471 PRELIMINARY; PRT; 8 AA.  
 AC 087471;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE H1FA (FRAGMENT).  
 GN H1FA.

OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EGAN (EIA);  
 RX MEDLINE=98389689; PubMed=9721313;  
 RA Mhlanga-Mutandadura T., Morlin G., Smith A.L., Eisenstark A.,  
 RA Golomb M.;  
 RT "Evolution of the major pilus gene cluster of haemophilus  
 RT influenzae.";  
 RL J. Bacteriol. 180:4693-4703(1998).  
 DR EMBL: AF071762; AAC35830.1; -;  
 FT NON\_TER  
 SQ SEQUENCE 8 AA; 876 MW; DDB44451A7272325 CRC64;

Query Match 28.3%; Score 13; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KILL 4  
 : 1 :  
 DB 2 KILL 4

RESULT 9  
 ID 002831 PRELIMINARY; PRT; 8 AA.  
 AC 002831;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE PRO ALPHA 1 TYPE III COLLAGEN PROTEIN (FRAGMENT).  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96377339; PubMed=8783186;  
 RA Metseranta M., Kujala O.M., Pelliniemi L., Osterman H., Aho H.,  
 RA Vuorio E.;  
 RT "Evidence for insufficient chondrocytic differentiation during repair  
 RT of full-thickness defects of articular cartilage.";  
 RL Matrix Biol. 15:39-47(1996).  
 DR EMBL: S83371; AAD14433.1; -;  
 RN Collagen.  
 FT NON\_TER  
 SQ SEQUENCE 8 AA; 1028 MW; B859C7272EA77371 CRC64;

Query Match 28.3%; Score 13; DB 6; Length 8;  
 Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
 Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 YW 7  
 : 1 :  
 DB 1 HW 2

RESULT 10  
 ID P92421 PRELIMINARY; PRT; 7 AA.  
 AC P92421; P92419;  
 DT 01-MAY-1997 (TREMBlrel. 03, Created)  
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last annotation update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11.  
 OS Psathyrostachys fragilis.  
 GN Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Trilicaceae; Poaceae; PACC clade;  
 OC NCBI\_TaxID=37723;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H4372, AND H917; TISSUE-LEAVES;  
 RA Petersen G., Seberg O.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: 277753; CAB01337.1; -;  
 DR EMBL: 277752; CAB01334.1; -;  
 KW Ribosomal protein; Chloroplast.  
 FT NON\_TER 1  
 SQ SEQUENCE 7 AA: 894 MW: 673414033277700 CRC64;

Query Match 26.1%; Score 12; DB 8; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 PK 2  
 DB 2 PK 3

RESULT 11  
 ID P92385 PRELIMINARY; PRT; 7 AA.  
 AC P92385; P92383;  
 DT 01-MAY-1997 (TREMUREL. 03, Created)  
 DT 01-MAY-1997 (TREMUREL. 03, Last sequence update)  
 DT 01-JAN-1998 (TREMUREL. 05, Last annotation update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11  
 OS Hordeum maritimum (Seaside barley).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Trilicaceae; Hordeum.  
 OC NCBI\_TaxID=4519;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H299, AND H801; TISSUE-LEAVES;  
 RA Petersen G., Seberg O.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: 277763; CAB01367.1; -;  
 DR EMBL: 277762; CAB01364.1; -;  
 KW Ribosomal protein; Chloroplast.  
 FT NON\_TER 1  
 SQ SEQUENCE 7 AA: 894 MW: 673414033277700 CRC64;

Query Match 26.1%; Score 12; DB 8; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 PK 2  
 DB 2 PK 3

RESULT 12  
 ID P92372 PRELIMINARY; PRT; 7 AA.  
 AC P92372;  
 DT 01-MAY-1997 (TREMUREL. 03, Created)  
 DT 01-MAY-1997 (TREMUREL. 03, Last sequence update)  
 DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11  
 OS Haynaldia villosa.  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Pacc clade; Andropogoneae; Haynaldia.  
 OC NCBI\_TaxID=40247;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H5561; TISSUE-LEAF;  
 RX MEDLINE-97271648; PubMed-9126564;  
 RA Petersen G., Seberg O.;  
 RT "Phylogenetic analysis of the Trilicaceae (Poaceae) based on rpoA  
 sequence data";  
 RL Mol. Phylogenet. Evol. 7:217-230(1997).  
 DR EMBL: 277741; CAB01301.1; -;  
 KW Chloroplast.  
 FT NON\_TER 1  
 SQ SEQUENCE 7 AA: 894 MW: 673414033277700 CRC64;

Query Match 26.1%; Score 12; DB 8; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 PK 2  
 DB 2 PK 3

RESULT 13  
 ID P92403 PRELIMINARY; PRT; 7 AA.  
 AC P92403;  
 DT 01-MAY-1997 (TREMUREL. 03, Created)  
 DT 01-MAY-1997 (TREMUREL. 03, Last sequence update)  
 DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11  
 OS Lolium perenne (Tall wheatgrass) (Aegropyrum elongatum).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Trilicaceae; Lolium.  
 OC NCBI\_TaxID=4588;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H6692; TISSUE-LEAF;  
 RX MEDLINE-97271648; PubMed-9126564;  
 RA Petersen G., Seberg O.;  
 RT "Phylogenetic analysis of the Trilicaceae (Poaceae) based on rpoA  
 sequence data";  
 RL Mol. Phylogenet. Evol. 7:217-230(1997).  
 DR EMBL: 277743; CAB01307.1; -;  
 KW Chloroplast.  
 FT NON\_TER 1  
 SQ SEQUENCE 7 AA: 894 MW: 673414033277700 CRC64;

Query Match 26.1%; Score 12; DB 8; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 PK 2  
 DB 2 PK 3

RESULT 14  
 ID P92425 PRELIMINARY; PRT; 7 AA.  
 AC P92425;  
 DT 01-MAY-1997 (TREMUREL. 03, Created)  
 DT 01-MAY-1997 (TREMUREL. 03, Last sequence update)  
 DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11.

OS Pseudoroegneria spicata.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Pseudoroegneria.  
 OX NCBI\_TaxID=4604;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H9082; TISSUE-LEAF;  
 RX MEDLINE=97271648; PubMed=9126564;  
 RA Petersen G., Seberg O.;  
 RM "Phylogenetic analysis of the Triticeae (Poaceae) based on rpoA  
 sequence data.";  
 RL Mol. Phylogenet. Evol. 7:217-230(1997).  
 DR EMBL: Z77744; CAB01310.1; -.  
 KW Chloroplast.  
 FT NON\_TER  
 SO SEQUENCE 7 AA; 894 MW; 673414033277700 CRC64;

Query Match 26.1%; Score 12; DB 8; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PK 2  
 ||  
 DB 2 PK 3

RESULT 15  
 ID P92387 PRELIMINARY; PRT; 7 AA.  
 AC P92387.  
 DT 01-MAY-1997 (TReMBLrel. 03, Created)  
 DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11.  
 OS Hemardia persica.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Hemardia.  
 OX NCBI\_TaxID=37678;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H3556; TISSUE-LEAF;  
 RX MEDLINE=97271648; PubMed=9126564;  
 RA Petersen G., Seberg O.;  
 RT "Phylogenetic analysis of the Triticeae (Poaceae) based on rpoA  
 sequence data.";  
 RL Mol. Phylogenet. Evol. 7:217-230(1997).  
 DR EMBL: Z77748; CAB01322.1; -.  
 KW Chloroplast.  
 FT NON\_TER  
 SO SEQUENCE 7 AA; 894 MW; 673414033277700 CRC64;

Query Match 26.1%; Score 12; DB 8; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PK 2  
 ||  
 DB 2 PK 3

Search completed: July 15, 2002, 13:42:04  
 Job time: 194 sec

10

11

12

13

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:27:15 ; Search time 28.83 seconds

(without alignments)  
30.822 Million cell updates/sec

Title: US-09-712-819A-4

Sequence: 1 PKLIYMA 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 69368

Minimum DB seq length: 0

Maximum DB seq length: 8

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	91.3	8	20	AAV40668 A2 derivative #10
2	42	91.3	8	21	AAV30007 Scaffold protein S
3	31	67.4	8	20	AAV40669 A2 derivative #11
4	31	67.4	8	21	AAV30008 Scaffold protein S
5	28	60.9	8	20	AAV40667 A2 derivative #9
6	28	60.9	8	21	AAV30006 Scaffold protein S
7	26	56.5	8	19	AAW37187 Human oncogenic pr
8	26	56.5	8	19	AAW37178 Human C-reactive p
9	26	56.5	8	20	AAW89743 Human C-reactive p
10	25	54.3	6	22	AAW82651 All-D peptide used
11	25	54.3	7	22	AAW65040 Human matrix metal

12	25	54.3	7	22	AAW35978
13	24	52.2	6	22	AAW56865
14	24	52.2	4	22	AAW82652
15	24	52.2	8	14	AAW32362
16	23	50.0	7	22	AAW44793
17	22	47.8	5	4	AAW30375
18	22	47.8	5	14	AAW41704
19	22	47.8	5	15	AAW47533
20	22	47.8	5	17	AAW13228
21	22	47.8	5	19	AAW87106
22	22	47.8	5	19	AAW87108
23	22	47.8	5	19	AAW70980
24	22	47.8	5	19	AAW40265
25	22	47.8	6	4	AAW30378
26	22	47.8	6	4	AAW30380
27	22	47.8	6	4	AAW30381
28	22	47.8	6	4	AAW30382
29	22	47.8	6	4	AAW30383
30	22	47.8	6	16	AAW93710
31	22	47.8	6	21	AAW85408
32	22	47.8	7	22	AAW44768
33	22	47.8	7	22	AAW46468
34	22	47.8	8	18	AAW17514
35	22	47.8	8	19	AAW20865
36	22	47.8	8	19	AAW4848
37	22	47.8	8	20	AAW10362
38	22	47.8	8	21	AAW23931
39	22	45.7	6	17	AAW05703
40	21	45.7	6	22	AAW9764
41	21	45.7	8	11	AAW03450
42	21	45.7	8	15	AAW63504
43	21	45.7	8	18	AAW4604
44	21	45.7	8	18	AAW17513
45	21	45.7	8	20	AAW97529

#### ALIGNMENTS

##### RESULT 1

AAV40668 AAV40668 standard; peptide: 8 AA.

AC AAV40668:

DT 01-DEC-1999 (first entry)

XX A2 derivative #10, beta strand of scaffold protein structure.

DE Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;

KW tumour; chemotherapeutic agent.

XX Synthetic.

OS Synthetic.

XX EP947582-A1.

PN 06-OCT-1999.

PD 31-MAR-1998; 98EP-0870065.

PE 31-MAR-1998; 98EP-0870065.

XX 31-MAR-1998; 98EP-0870065.

XX (INNO-) INNOGENETICS NV.

XX Desmet J, Hufon S, Hoogenboom H, Sablon E;

XX WPT; 1999-542958/46.

XX New scaffold protein, useful for stabilizing antigens used as vaccines

XX Disclosure; Page 5; 105pp; English.

Collagenase cleave  
Human SNP related  
All-D peptide used  
Sample peptide to  
H11 binding site c  
Peptide having pit  
GHRP-6 (Growth hor  
GHRP-6 analog. Sy  
Growth hormone rel  
Peptide determined  
Peptide determined  
Growth hormone rel  
K. oxytoca R-spect  
Peptide having pit  
Peptide having pit  
Peptide having pit  
Peptide having pit  
Peptide having pit  
CycloI-Tyr-trp-Ala  
Human HEGC protein  
H11 binding site c  
H11 binding site c  
Protein kinase C r  
Human presentin I  
Cyclomaltodextrin  
T cell epitope/MHC  
Somatostatin pepti  
Peptide motif need  
Rhesus D antibody  
Accessory molecu d  
TMD5 region amino  
Anti-fungal peptid  
Protein Kinase C r  
Antigenic site of

CC Sequences AAY40659-Y40679 are functionally equivalent derivatives of the  
CC A2 peptide (AAY40604) which forms part of a scaffold protein. A2 is a  
CC beta strand peptide which forms part of a beta sheet. Peptides  
CC (AAY40601-Y40609) together form a single-chain scaffold protein which  
CC contains at least 1 disulfide bond, contains less than 10% alpha helix  
CC and contains at least 6 beta-strands. The scaffold protein is constructed  
CC of beta strands SI-S6, and may also include beta strands AI-A3, or any  
CC functionally equivalent derivative of these sequences. The beta strands  
CC form two beta sheets SI/S4/S3 and S6/S5/S2 with each strand connected to  
CC the next by hydrogen bonds, which generate a beta sandwich architecture.  
CC If the additional beta strands AI-A3 are included in the structure the  
CC scaffold is constructed of two beta sheets, with the structures  
CC AI/SI/S4/S3 and S6/S5/S2/A2/A3. The beta strands are connected to each  
CC other via amino acid loops, where at least one of the loops binds to a  
CC receptor or antigen. The scaffold protein is used to stabilize antigens  
CC or whole proteins such as receptors, or their fragments. It may be used  
CC to bind two separate molecules. For example, one surface of the scaffold  
CC may be bound to a protein which binds to a tumour antigen. This will  
CC target the complex to tumour cells. Another surface may be bound to a  
CC cytotoxic molecule or an autoimmune antibody which may then kill the  
CC tumour cells. Therefore the scaffold protein may be used to target  
CC chemotherapeutic agents to specific cells. It may also be used to  
CC stabilize individual peptides in a peptide library and may be used in  
CC diagnostic techniques, and to stabilize antigens used as vaccines.

CC Sequence 8 AA:

SO

Query Match 91.3%; Score 42; DB 20; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLLIYW 7  
| | | | |  
DB 2 PKLLIYW 8

RESULT 2  
AAB30007  
ID AAB30007 standard; Peptide: 8 AA.  
XX  
AC AAB30007;  
XX  
DT 09-FEB-2001 (first entry)  
XX  
DE Scaffold protein SCA S2 peptide seq ID NO: 68.  
XX  
KM Human; CTIA-4; scaffold protein; antigen-binding; receptor-binding;  
KM SCA domain; cancer; thrombosis; osteoporosis, Rheumatoid arthritis;  
KM diabetic retinopathy; atherosclerosis.  
XX  
OS Synthetic.  
XX  
PN WC200060070-A1.  
XX  
PD 12-OCT-2000.  
XX  
PF 01-APR-1999; 99WO-EP02283.  
XX  
PR 01-APR-1999; 99WO-EP02283.  
XX  
PA (INNO-) INNOGENETICS NV.  
XX  
PI Desmet J, Hufton S, Hoogenboom H, Sablon E;  
DR WPI: 2000-665002/64.  
XX  
XX Scaffold composed of single-chain polypeptide having beta sandwich  
PT architecture carrying new and randomized peptide sequences useful as  
PT supporting framework and carrying antigen- or receptor binding  
PT fragments -  
XX  
PS Disclosure; Page 14: 68pp; English.

XX  
CC The present invention is concerned with producing scaffold proteins  
CC based upon the human CTIA-4 SCA domain. These scaffold proteins can be  
CC used as a scaffold to bind antigen- or receptor-binding fragments. These  
CC can be used in the treatment of diseases such as cancer,  
CC atherosclerosis, thrombosis, osteoporosis, Rheumatoid arthritis and  
CC diabetic retinopathy. Sequences AAY29930-B29939 were used in the  
CC production of the proteins of the invention.

CC Sequence 8 AA:

SO

Query Match 91.3%; Score 42; DB 21; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLLIYW 7  
| | | | |  
DB 2 PKLLIYW 8

RESULT 3  
AAY40669  
ID AAY40669 standard; peptide: 8 AA.  
XX  
AC AAY40669;  
XX  
DT 01-DEC-1999 (first entry)  
XX  
DE A2 derivative #11, beta strand of scaffold protein structure.  
XX  
DE Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;  
XX  
KM tumour; chemotherapeutic agent.  
XX  
OS Synthetic.  
XX  
PN EP947582-A1.  
XX  
PD 06-OCT-1999.  
XX  
PE 31-MAR-1998; 98EP-0870065.  
XX  
PR 31-MAR-1998; 98EP-0870065.  
XX  
PA (INNO-) INNOGENETICS NV.  
XX  
PI Desmet J, Hufton S, Hoogenboom H, Sablon E;  
DR WPI: 1999-542958/46.  
XX  
XX New scaffold protein, useful for stabilizing antigens used as vaccines  
PT -  
PT Disclosure; Page 5; 105pp; English.  
XX  
XX Sequences AAY40659-Y40679 are functionally equivalent derivatives of the  
CC A2 peptide (AAY40604) which forms part of a scaffold protein. A2 is a  
CC beta strand peptide which forms part of a beta sheet. Peptides  
CC (AAY40601-Y40609) together form a single-chain scaffold protein which  
CC contains at least 1 disulfide bond, contains less than 10% alpha helix  
CC and contains at least 6 beta-strands. The scaffold protein is constructed  
CC of beta strands SI-S6, and may also include beta strands AI-A3, or any  
CC functionally equivalent derivative of these sequences. The beta strands  
CC form two beta sheets SI/S4/S3 and S6/S5/S2 with each strand connected to  
CC the next by hydrogen bonds, which generate a beta sandwich architecture.  
CC If the additional beta strands AI-A3 are included in the structure the  
CC scaffold is constructed of two beta sheets, with the structures  
CC AI/SI/S4/S3 and S6/S5/S2/A2/A3. The beta strands are connected to each  
CC other via amino acid loops, where at least one of the loops binds to a  
CC receptor or antigen. The scaffold protein is used to stabilize antigens  
CC or whole proteins such as receptors, or their fragments. It may be used  
CC to bind two separate molecules. For example, one surface of the scaffold  
CC may be bound to a protein which binds to a tumour antigen. This will



target the complex to tumour cells. Another surface may be bound to a cytotoxic molecule or an autoimmune antibody which may then kill the tumour cells. Therefore the scaffold protein may be used to target chemotherapeutic agents to specific cells. It may also be used to stabilize individual peptides in a peptide library and may be used in diagnostic techniques, and to stabilize antigens used as vaccines.

Sequence 8 AA:

Query Match 67.4%; Score 31; DB 20; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLITY 6  
1:|||||  
Db 2 PKLITY 7

#### RESULT 4

AAB30008  
ID AAB30008 standard; Peptide; 8 AA.

AC AAB30008;

DT 09-FEB-2001 (first entry)

Scaffold protein SCA S2 peptide SEQ ID NO: 69.

Human; CTLA-4; scaffold protein; antigen-binding; receptor-binding;  
SCA domain; cancer; thrombosis; osteoporosis; rheumatoid arthritis;  
diabetic retinopathy; atherosclerosis.

Synthetic.

WO200060070-A1.

12-OCT-2000.

01-APR-1999; 99WO-EP02283.

01-APR-1999; 99WO-EP02283.

(INNO-) INNOGENETICS NV.

Desmet J, Hufon S, Hoogenboom H, Sablon E;

WPI: 2000-665002/64.

Scaffold composed of single-chain polypeptide having beta sandwich architecture carrying new and randomized peptide sequences useful as supporting framework and carrying antigen- or receptor binding fragments

Disclosure; Page 14; 68pp; English.

The present invention is concerned with producing scaffold proteins based upon the human CTLA-4 SCA domain. These scaffold proteins can be used as a scaffold to bind antigen- or receptor-binding fragments. These can be used in the treatment of diseases such as cancer, atherosclerosis, thrombosis, osteoporosis, rheumatoid arthritis and diabetic retinopathy. Sequences AAB39930-B29939 were used in the production of the proteins of the invention.

Sequence 8 AA:

Query Match 67.4%; Score 31; DB 21; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLITY 6  
1:|||||

Db 2 PKLITY 7

RESULT 5  
AAV40667  
ID AAV40667 standard; peptide; 8 AA.

AC AAV40667;

DT 01-DEC-1999 (first entry)

A2 derivative #9, beta strand of scaffold protein structure.

Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;  
tumour; chemotherapeutic agent.

Synthetic.

EP947582-A1.

06-OCT-1999.

31-MAR-1998; 98EP-0870065.

31-MAR-1998; 98EP-0870065.

(INNO-) INNOGENETICS NV.

Desmet J, Hufon S, Hoogenboom H, Sablon E;

WPI: 1999-542958/46.

New scaffold protein, useful for stabilizing antigens used as vaccines

Disclosure; Page 5; 105pp; English.

Sequences AAV40659-Y40679 are functionally equivalent derivatives of the A2 peptide (AAV40604) which forms part of a scaffold protein. A2 is a beta strand peptide which forms part of a beta sheet. Peptides (AAV40601-Y40609) together form a single-chain scaffold protein which contains at least 1 disulfide bond, contains less than 10% alpha helix and contains at least 6 beta-strands. The scaffold protein is constructed of beta strands S1-S6, and may also include beta strands A1-A3, or any functionally equivalent derivative of these sequences. The beta strands form two beta sheets S1/S4/S3 and S6/S5/S2 with each strand connected to the next by hydrogen bonds, which generate a beta sandwich architecture. If the additional beta strands A1-A3 are included in the structure the scaffold is constructed of two beta sheets, with the structures A1/S1/S4/S3 and S6/S5/S2/A2/A3. The beta strands are connected to each other via amino acid loops, where at least one of the loops binds to a receptor or antigen. The scaffold protein is used to stabilize antigens or whole proteins such as receptors, or their fragments. It may be used to bind two separate molecules. For example, one surface of the scaffold may be bound to a protein which binds to a tumour antigen. This will target the complex to tumour cells. Another surface may be bound to a cytotoxic molecule or an autoimmune antibody which may then kill the tumour cells. Therefore the scaffold protein may be used to target chemotherapeutic agents to specific cells. It may also be used to stabilize individual peptides in a peptide library and may be used in diagnostic techniques, and to stabilize antigens used as vaccines.

Sequence 8 AA:

Query Match 60.9%; Score 28; DB 20; Length 8;  
Best Local Similarity 83.3%; Pred. No. 6.4e+05;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLITY 6  
1:|||||  
Db 2 PKLITY 7

RESULT 6  
AAB30006 standard; Peptide: 8 AA.  
XX ID AAB30006; Peptide: 8 AA.  
XX AC AAB30006;  
XX XX  
XX 09-FEB-2001 (first entry)  
XX XX  
DE Scaffold protein SCA S2 peptide seq ID NO: 67.  
XX XX  
KW Human; CTLA-4; scaffold protein; antigen-binding; receptor-binding;  
KW SCA domain; cancer; thrombosis; osteoporosis; rheumatoid arthritis;  
KW diabetic retinopathy; atherosclerosis.  
XX OS  
XX Synthetic.  
XX PN W0200060070-A1.  
XX PD 12-OCT-2000.  
XX PF 01-APR-1999; 99WO-EP02283.  
XX PR 01-APR-1999; 99WO-EP02283.  
XX PA (INNO-) INNOGENETICS NV.  
XX PI Desmet J, Hufton S, Hoogenboom H, Sablon E;  
XX WPI: 2000-665002/64.  
XX DR  
XX PT Scaffold composed of single-chain polypeptide having beta sandwich  
PT architecture carrying new and randomized peptide sequences useful as  
PT supporting framework and carrying antigen- or receptor binding  
PT fragments -  
XX PS  
XX PS Disclosure: Page 14; 68pp: English.  
XX CC  
XX CC The present invention is concerned with producing scaffold proteins  
CC based upon the human CTLA-4 SCA domain. These scaffold proteins can be  
CC used as a scaffold to bind antigen- or receptor-binding fragments. These  
CC can be used in the treatment of diseases such as cancer,  
CC atherosclerosis, thrombosis, osteoporosis, rheumatoid arthritis and  
CC diabetic retinopathy. Sequences AAB2930-B29939 were used in the  
CC production of the proteins of the invention.  
XX CC  
SQ Sequence 8 AA:  
  
Query Match 60.9%; Score 28; DB 21; Length 8;  
Best Local Similarity 83.3%; Pred. No. 6.4e+05;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 1 PKLLTY 6  
I : I I I I  
Db 2 pLLTY 7  
  
RESULT 7  
AAW37187  
ID AAW37187 standard; peptide: 8 AA.  
XX AC AAW37187;  
XX XX  
XX 20-JUL-1998 (first entry)  
XX XX  
DE Human oncogenic protein MDM2 binding N-acetylated peptide derivative 6.  
XX XX  
KW MDM2; oncogenic protein; p53; human; inhibition; interaction; cancer;  
KW tumour; diagnosis; binding; viral infection.  
XX OS  
XX Synthetic.  
OS Homo sapiens.

XX Key Location/Qualifiers  
FH Modified-site 1  
FT Modified-site /note="N-terminal acetyl"  
FT Modified-site 8  
FT /note="C-terminal amide"  
XX  
XX M09801467-A2.  
XX PD 15-JAN-1998.  
XX PF 04-JUL-1997; 97WO-EP03549.  
XX PR 07-APR-1997; 97GB-0007041.  
XX PR 05-JUL-1996; 96GB-0014197.  
XX PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.  
XX PA (NOVS ) NOVARTIS AG.  
XX PI Boettger A, Boettger V, Chene P, Furet P, Garcia-echaverria C;  
XX PI Hochkeppel H, Lane D, Pickstey S;  
XX DR WPI: 1998-100996/09.  
XX PT Compounds binding to MDM2 protein and inhibit its interaction with  
PT p53 - useful in, e.g. diagnosis and treatment of cancer and viral  
PT infections and identifying binding agents  
XX PS  
XX PS Example 1: Page 19; 45pp: English.  
XX CC  
XX CC This is a N-acetylated peptide derivative capable of binding to a human  
CC oncogenic protein MDM2. The MDM2 binding peptides can specifically  
CC inhibit or block the binding of MDM2 to the human p53 protein, in vitro  
CC or in vivo. Inhibiting the interaction between the p53 and MDM2 can  
CC induce growth arrest or apoptosis in tumour cells comprising a wild-type  
CC p53 and non-elevated levels of MDM2. The peptides may be used to identify  
CC molecules that bind to MDM2 and to identify and design inhibitors of  
CC MDM2/p53 binding. They may also be used to purify binding partners  
CC especially MDM2, diagnose disease by measuring levels of MDM2 in blood  
CC of cancer and leukaemia patients and for treatment or prevention of  
CC disease involving p53/MDM2 interactions, especially tumours and viral  
CC infections. The peptides can be administered nasally, rectally, orally  
CC or by injection. By interfering with MDM2/p53 interaction, the peptides  
CC can activate p53 function and accumulation in normal cells. The peptides  
CC which mimic the MDM2 binding site in p53, have a significantly greater  
CC blocking activity compared with wild-type p53.  
XX CC  
SQ Sequence 8 AA:  
  
Query Match 56.5%; Score 26; DB 19; Length 8;  
Best Local Similarity 42.9%; Pred. No. 6.4e+05;  
Matches 3; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 1 PKLLTY 7  
I : I I I I  
Db 1 pLLTY 7  
  
RESULT 8  
AAW37178  
ID AAW37178 standard; peptide: 8 AA.  
XX AC AAW37178;  
XX XX  
XX 20-JUL-1998 (first entry)  
XX XX  
DE Human oncogenic protein MDM2 binding peptide derivative 6.  
XX XX  
KW MDM2; oncogenic protein; p53; human; inhibition; interaction; cancer;  
KW tumour; diagnosis; binding; viral infection.  
XX OS  
XX Synthetic.  
OS Homo sapiens.

```

OS Homo sapiens.
XX
XX MO9801467-A2.
XX
XX 15-JAN-1998.
XX
XX 04-JUL-1997; 97MO-EP03549.
XX
XX 07-APR-1997; 97GB-0007041.
XX
XX 05-JUL-1996; 96GB-0014197.
XX
XX (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
XX (NOVS ) NOVARTIS AG.
XX
XX Boettiger A, Boettiger V, Chene P, Furet P, Garcia-echeverria C;
PI Hockeppel H, Lane D, Picklesley S;
XX WPI: 1998-100996/09.
XX
XX Compounds binding to MDM2 protein and inhibit its interaction with
PT p53 - useful in, e.g. diagnosis and treatment of cancer and viral
PT infections and identifying binding agents
XX
XX Claim 9; Page 42; 45pp; English.
XX
XX This peptide is capable of binding to an oncogenic protein MDM2
CC (especially human MDM2). The MDM2 binding peptides can specifically
CC inhibit or block the binding of MDM2 to the human p53 protein, in vitro
CC or in vivo. Inhibiting the interaction between the p53 and MDM2 can
CC induce growth arrest or apoptosis in tumour cells comprising a wild-type
CC p53 and non-elevated levels of MDM2. The peptides may be used to identify
CC molecules that bind to MDM2 and to identify and design inhibitors of
CC MDM2/p53 binding. They may also be used to purify binding partners
CC especially MDM2, diagnose disease by measuring levels of MDM2 in blood
CC of cancer and leukaemia patients and for treatment or prevention of
CC disease involving p53/MDM2 interactions, especially tumours and viral
CC infections. The peptides can be administered nasally, rectally, orally
CC or by injection. By interfering with MDM2/p53 interaction, the peptides
CC can activate p53 function and accumulation in normal cells. The peptides
CC which mimic the MDM2 binding site in p53, have a significantly greater
CC blocking activity compared with wild-type p53.
XX
XX Sequence 8 AA:
SQ
Query Match 56.5%; Score 26; DB 19; Length 8;
Best Local Similarity 42.9%; Pred. No. 6.4e+05;
Matches 3; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 PRLIYW 7
DB 1 pfmdyw 7
RESULT 9
AAB89743
ID AAB89743 standard; peptide; 8 AA.
XX
XX AAB89743;
XX
XX 20-APR-1999 (first entry)
XX
XX Human C-reactive protein (CRP)-derived peptide 15 (residues 62-69).
DE
XX C-reactive protein; CRP; hLE; inhibit; enzymatic; anti-inflammatory;
XX human Leukocyte Elastase; human Cathepsin G; Rheumatoid arthritis; hCG;
XX pulmonary emphysema; asthma; acute respiratory distress syndrome;
XX cystic fibrosis; bronchitis; connective tissue; human.
XX
XX Synthetic.
OS Homo sapiens.
XX
XX MO9900418-A1.

```

```

XX
XX 07-JAN-1999.
XX
XX 29-JUN-1998; 98MO-IL00302.
XX
XX 29-JUN-1997; 97IL-0121191.
XX
XX (YEDA ) YEDA RES & DEV CO LTD.
XX
XX Fridkin M, Yavin EJ;
XX WPI: 1999-095681/08.
XX
XX New peptide(s) derived from human C-reactive protein - useful for
PT inhibiting, in vitro, enzymatic activity of human leukocyte elastase
PT and/or human cathepsin G in the treatment of chronic inflammatory
PT conditions
XX
XX Example 1; Page 15; 33pp; English.
XX
XX The invention relates to anti-inflammatory peptides derived from human
CC C-reactive protein (CRP) that are capable of inhibiting in vitro
CC enzymatic activity of human Leukocyte Elastase (hLE) and/or human
CC Cathepsin G (hCG). The CRP peptide derivatives are useful for treating
CC chronic inflammatory conditions, especially rheumatoid arthritis,
CC pulmonary emphysema, cystic fibrosis, bronchitis, acute respiratory
CC distress syndrome, asthma and other chronic inflammatory tissue
CC destructive conditions. The CRP peptide derivatives significantly
CC inhibit the enzymatic activities of hLE and hCG enzymes (released from
CC neutrophils), which prevents permanent connective tissue damage
CC associated with inflammatory conditions. Sequences AAB89733-743 and
CC AA30578 -581 represent peptide fragments derived from human CRP. These
CC are comparison peptides that were used along with the claimed
CC CRP-derived peptides of the invention (AAB89722-732).
XX
XX Sequence 8 AA:
SQ
Query Match 56.5%; Score 26; DB 20; Length 8;
Best Local Similarity 42.9%; Pred. No. 6.4e+05;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 2 KLIIYWA 8
DB 1 eilfws 7
RESULT 10
AAB82651
ID AAB82651 standard; Peptide; 6 AA.
XX
XX AAB82651;
XX
XX 02-OCT-2001 (first entry)
XX
XX All-D peptide used in Alzheimer's disease vaccine.
DE
XX Alzheimer's disease; amyloidosis; amyloid-related disease;
XX vaccine; therapy; antigen.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Misc-difference 1..6
FT /note= "all D-form residues"
FT Modified-site 6
FT /note= "C-terminal amide"
XX
XX MO200139796-A2.
XX
XX 07-JUN-2001.
XX
XX 29-NOV-2000; 2000MO-CA01413.

```

XX 29-NOV-1999; 99US-0168594.  
 PR 28-NOV-2000; 2000US-0724842.  
 XX (NEUR-) NEUROCHEM INC.  
 PA  
 PI Chalfour R, Hebert L, Kong X, Gervais F;  
 XX WPI: 2001-441458/47.  
 DR  
 XX Preventing/treating amyloid-related disease, especially Alzheimer's  
 PT disease, comprises administering antigenic all-D peptide, e.g. as  
 PT vaccine, which elicits production of antibodies to prevent  
 PT fibrillogenesis and associated cellular toxicity.  
 XX  
 XX Disclosure: Page 12: 31pp; English.  
 PS  
 XX The present sequence is that of an all-D peptide suitable for  
 CC use for preparing vaccines for preventing or treating Alzheimer's  
 CC disease and other amyloid related disorders in humans. It is based  
 CC on a portion of amyloid-beta peptide (see AAB82622). Vaccines  
 CC are produced using 'non-self' peptides synthesized from the  
 CC unnatural D-configuration amino acids to avoid the drawbacks of  
 CC 'self' proteins. The all-D peptides need not be aggregated to be  
 CC operative or immunogenic. They preferably interact with at  
 CC least 1 region of an amyloid protein, e.g. the beta-sheet region  
 CC or GAG-binding site region, the amyloid-beta peptide, or their  
 CC immunogenic fragments, protein conjugates, immunogenic derivative  
 CC peptides and immunogenic peptidomimetics. Examples include all-D  
 CC peptides corresponding to residues 1-42, 1-40, 1-35, 1-28, 1-7,  
 CC 10-16, 16-21 and 36-42 of the amyloid-beta peptide and the all-D  
 CC derivative peptides given in AAB82623-64. The vaccine elicits a  
 CC preferential TH-2 or TH-1 response, preventing fibrillogenesis and  
 CC associated cellular toxicity. The amyloid related diseases may be  
 CC localised amyloidosis, e.g. diabetes type II, neurodegenerative  
 CC diseases, e.g. bovine spongiform encephalitis, Creutzfeldt-Jakob  
 CC disease, scrapie, cerebral amyloid angiopathy, and prion protein  
 CC related disorders, or systemic amyloidosis associated with chronic  
 CC infection (e.g. tuberculosis) or chronic inflammation (e.g.  
 CC rheumatoid arthritis), familial Mediterranean fever (FMF) and  
 CC systemic amyloidosis found in long-term haemodialysis patients.  
 XX  
 SQ Sequence 6 AA:  
 Query Match 54.3%; Score 25; DB 22; Length 6;  
 Best Local Similarity 60.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 OY 4 LIYMA 8  
 Db 2 LAYMA 6  
 RESULT 11  
 AAG65040  
 ID AAG65040 standard; peptide: 7 AA.  
 XX  
 AC AAG65040:  
 XX  
 DT 09-OCT-2001 (first entry)  
 XX  
 XX Human matrix metalloproteinase MMP-8 substrate peptide.  
 DE  
 XX Chemical sensor system; microjet; indicator chemistry; ligand analysis;  
 KW biomedicine; environmental monitoring; biowarfare.  
 XX  
 OS Synthetic.  
 XX  
 FM Key Location/Qualifiers  
 FT Modified-site 1  
 FT /label= OTHER  
 FT /note= "modified by Dnp"

FT Modified-site 7  
 FT /label= OTHER  
 FT /note= "C-terminal amide"  
 XX  
 XX W0200157494-A2.  
 XX  
 XX 09-AUG-2001.  
 PD  
 XX  
 XX 17-JAN-2001; 2001WO-US01553.  
 XX  
 XX 20-JAN-2000; 2000US-0177105.  
 PR 09-NOV-2000; 2000US-0709047.  
 XX  
 XX (RDCG) UNIV CALIFORNIA.  
 PA  
 XX Brown SB, Colston BW, Langry K, Milancovich FP, Simon J, Cox WR,  
 PI Hayes DJ;  
 XX WPI: 2001-488913/53.  
 DR  
 XX  
 XX Producing chemical sensors useful for biomedical, environmental,  
 PT occupational safety, process control, biowarfare applications, by  
 PT printing indicator chemistries on optically accessible surface by  
 PT microjet technology.  
 XX  
 XX Disclosure: Page 13: 31pp; English.  
 PS  
 XX The present invention relates to a method of producing a chemical sensor,  
 CC involving the use of microjet technology to print one or more indicator  
 CC chemistries on an optically accessible surface. This is useful for  
 CC producing a chemical sensor useful for detecting and/or analysing a  
 CC sample in a fluid or airborne medium, for monitoring hazardous materials  
 CC such as heavy metals, hydrocarbons and chlorinated hydrocarbons in both  
 CC ground water and soil of contaminated sites, for making accurate  
 CC dosimetry measurements of hazardous materials, such as carcinogens or  
 CC mutagens present in hostile or potentially hostile environments. The  
 CC sensor systems can be implemented in assembly line type configurations  
 CC for quality and process control type applications, e.g. measurements of  
 CC gases emitted from fruits and vegetables and detection of contaminants in  
 CC soft drink or bottled water solutions. The systems are used for detection  
 CC of airborne or water-based chemical and biowarfare agents such  
 CC as anthrax and are suitable for measuring multiple constituents in a  
 CC small sample volume. The present sequence is a substrate for a matrix  
 CC metalloproteinase described in the exemplification of the invention.  
 XX  
 SQ Sequence 7 AA:  
 Query Match 54.3%; Score 25; DB 22; Length 7;  
 Best Local Similarity 80.0%; Pred. No. 6.4e+05;  
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 4 LIYMA 8  
 Db 2 LAYMA 6  
 RESULT 12  
 AAB35978  
 ID AAB35978 standard; peptide: 7 AA.  
 XX  
 AC AAB35978:  
 XX  
 DT 01-MAR-2001 (first entry)  
 XX  
 XX Collagenase cleavable peptide #7.  
 DE  
 XX Polymetric drug conjugate; enzymatically cleavable linker; candant;  
 KW antiinflammatory; cytostatic; hepatotropic; neuroprotective; cancer;  
 KW antibacterial; nephrotoxic; immunomodulatory; neoplastic disease;  
 KW chronic inflammatory disease; acute inflammatory disease;  
 KW cardiac disease; renal disease; liver disease; lung disease;  
 KW neurological disease; musculoskeletal disease; immunological disorder;

XX microbial infection.  
OS Synthetic.  
XX  
XX WO200064486-A2.  
XX  
XX  
XX 02-NOV-2000.  
XX  
XX 28-APR-2000; 2000WO-US11670.  
XX  
XX 28-APR-1999; 99US-0131404.  
XX  
XX 02-NOV-1999; 99US-0163090.  
XX  
XX (VERI-) VERITAS MEDICAL TECHNOLOGIES INC.  
XX  
XX Pachence JM, Belinka BA, Ramani T;  
XX  
XX WPI; 2001-031659/04.  
XX  
XX Polymeric drug conjugate, for treating diseases associated with organs  
XX PT e.g. liver or heart, has biologically active agents linked to regular  
XX PT repeating linear or branched co-polymers by enzymatically cleavable  
XX PT marker.  
XX  
XX Claim 43; Page 94; 100pp; English.  
XX  
XX This invention relates to a polymeric drug conjugate. The drug conjugate  
XX CC comprises biologically active agents conjugated via an enzymatically  
XX CC cleavable linker to a regular repeating linear unit comprising a water  
XX CC soluble polymer segment and a multifunctional chemical moiety, or to a  
XX CC branched polymer comprising two or more water soluble polymer segments  
XX CC each bound to a common multifunctional chemical moiety. The polymeric  
XX CC drug conjugate has antiinflammatory; cytostatic; cardiant; hepatotropic;  
XX CC neuroprotective; antibacterial; nephrotropic; and immunomodulatory  
XX CC activity. The drug conjugate is useful for alleviating a pathological  
XX CC condition such as neoplastic diseases, chronic inflammatory diseases  
XX CC acute inflammatory diseases, cardiac diseases, renal diseases, liver  
XX CC diseases, lung diseases, neurological diseases, musculoskeletal diseases  
XX CC and immunological disorders and various microbial infections by  
XX CC modulating immunological or hormonal function. The present sequence  
XX CC represents a peptide which can be used as the enzymatically cleavable  
XX CC linker in the drug conjugate of the invention.  
XX  
XX Sequence 7 AA:  
SQ  
Query Match 54.3%; Score 25; DB 22; Length 7;  
Best Local Similarity 80.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 4 LTYMA 8  
| | | |  
Db 2 LTYMA 6  
RESULT 13  
ABBS6865  
ID ABBS6865 standard; Peptide; 4 AA.  
ABBS6865;  
05-MAR-2002 (first entry)  
XX  
XX Human SNP related amino acid sequence SEQ ID NO:1430.  
XX  
XX DE  
XX Human: single nucleotide polymorphism; SNP; polymorphism; cytostatic;  
XX KM immunosuppressive; antiinflammatory; neuroprotective; antimicrobial;  
XX KM autoimmune disease; inflammation; cancer; nervous system disease;  
XX KM infection; polymorphic protein.  
XX  
XX Homo sapiens.  
XX  
XX CS  
XX WO200138586-A2.

PD	31-MAY-2001.
XX	
PF	22-NOV-2000; 2000MO-US32311.
XX	
PR	24-NOV-1999; 99US-0167383.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
P1	Shinkets RA, Leach M;
DR	WPI; 2001-355949/37.
PT	Isolated human nucleic acids comprising one or more single nucleotide polymorphisms, useful for treating a subject suffering from a pathology, e.g. autoimmune diseases, ascribed to the presence of a sequence polymorphism -
PS	Claim 1; Page 665; 674pp; English.
CC	AB00010 to AB01104 represent human nucleic acid oligonucleotides comprising one or more single nucleotide polymorphisms (SNPs) ABB56531 to ABB56903 represent human peptides encoded by some of the SNPs Oligonucleotides. The sequences from the present invention can have immunosuppressive, cytostatic, antiinflammatory, neuroprotective and antimicrobial activities. Nucleic acids, polypeptides, oligonucleotides and antibodies from the present invention can be used for treating a subject suffering from, at risk for, or suspected of, suffering from a pathology ascribed to the presence of a sequence polymorphism. The pathology may be autoimmune diseases, inflammation, cancer, diseases of the nervous system, and infection by pathogenic microorganisms. The SNPs are also useful for determining which forms of a characterized polymorphism are present in individuals. The antibodies may be used in the detection, quantitation and/or cellular or tissue localisation of a polymorphic protein (e.g., for use in measuring levels of the polymorphic protein within appropriate physiological samples).
SQ	Sequence 4 AA:
Query Match	52.2%; Score 24; DB 22; Length 4;
Best Local Similarity	75.0%; Pred. No. 6.4e+05;
Matches 3; Conservative	1; Mismatches 0; Indels 0; Gaps 0;
DY	4 LIYW 7   :     1 LIYW 4
Db	
RESULT 14	
AAB82652	
ID	AAB82652 standard; Peptide; 6 AA.
XX	
AC	AAB82652;
DT	
02-OCT-2001	(first entry)
XX	
DE	All-D peptide used in Alzheimer's disease vaccine.
XX	
KW	Alzheimer's disease; amyloidosis; amyloid-related disease; vaccine; therapy; antigen.
XX	
OS	Synthetic.
XX	
FH	Key location/Qualifiers
FT	Misc-difference 1..6
FT	/note= "all D-form residues"
FT	Modified-site 6
XX	/note= "C-terminal amide"
XX	
WO200139796-A2.	
XX	
DN	
07-JUN-2001.	
DD	

```

XX 29-NOV-2000; 2000MO-CM01413.
PF
XX
XX 29-NOV-1999; 99US-0168594.
PR
XX 28-NOV-2000; 2000US-0724842.
PR
XX (NEUR-) NEUROCHEM INC.
PA
XX
XX
XX Chailfour R, Hebert L, Kong X, Gervais F;
PI
XX
XX WPI; 2001-441458/47.
DR
XX
XX Preventing/treating amyloid-related disease, especially Alzheimer's
PT disease, comprises administering antigenic all-D peptide, e.g. as
PT vaccine, which elicits production of antibodies to prevent
PT fibrillogenesis and associated cellular toxicity.
PT
XX
XX Disclosure: Page 12; 31pp; English.
PS
XX
XX The present sequence is that of an all-D peptide suitable for
CC use for preparing vaccines for preventing or treating Alzheimer's
CC disease and other amyloid related disorders in humans. It is based
CC on a portion of amyloid-beta peptide (see AAB82622). Vaccines
CC are produced using 'non-self' peptides synthesised from the
CC unnatural D-configuration amino acids to avoid the drawbacks of
CC 'self' proteins. The all-D peptides need not be aggregated to be
CC operative or immunogenic. They preferably interact with at
CC least 1 region of an amyloid protein, e.g. the beta-sheet region
CC or GAG-binding site region, the amyloid-beta peptide, or their
CC immunogenic fragments, protein conjugates, immunogenic derivative
CC peptides and immunogenic peptidomimetics. Examples include all-D
CC peptides corresponding to residues 1-42, 1-40, 1-35, 1-28, 1-7,
CC 10-16, 16-21 and 36-42 of the amyloid-beta peptide and the all-D
CC derivative peptides given in AAB82623-64. The vaccine elicits a
CC preferential TH-2 or TH-1 response, preventing fibrillogenesis and
CC associated cellular toxicity. The amyloid related diseases may be
CC localised amyloidosis, e.g. diabetes type II, neurodegenerative
CC diseases, e.g. bovine spongiform encephalitis, Creutzfeldt-Jakob
CC disease, scrapie, cerebral amyloid angiopathy, and prion protein
CC related disorders, or systemic amyloidosis associated with chronic
CC infection (e.g. tuberculosis) or chronic inflammation (e.g.
CC rheumatoid arthritis), familial Mediterranean fever (FMF) and
CC systemic amyloidosis found in long-term haemodialysis patients.
CC
XX
SQ Sequence 6 AA;

Query Match 52.2%; Score 24; DB 22; Length 6;
Best Local Similarity 60.0%; Pred. No. 6.4e+05;
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLYMA 8
   1::||
   2 LYYMA 6

DB

RESULT 15
AAR32362
ID AAR32362 standard; peptide: 8 AA.
XX
XX AAR32362:
AC
XX 17-JUN-1993 (first entry)
DT
XX
XX Sample peptide to demonstrate a new sequencing method.
DE
XX
XX Sequential degradation; perfluoroalkanoic acid; anhydride.
KM
XX
XX Synthetic.
OS
XX
XX EP5329604-A.
PN
XX
XX 03-MAR-1993.
PD

```

```

XX 26-AUG-1992; 92EP-0114531.
PF
XX
XX 28-AUG-1991; 91JP-0217437.
PR
XX 15-NOV-1991; 91JP-0300818.
PR
XX (DASE ) SEIKO INSTR INC.
PA
XX
XX
XX Satake K, Takamoto K, Tsugita A, Uchida T;
PI
XX
XX WPI; 1993-068671/09.
DR
XX
XX Aminoacid sequencing of proteins or peptide(s) - by sequential
PT degradation with per:fluoro-alkanoic acid or anhydride
PT
XX
XX Example 4; Page 10; 38pp; English.
PS
XX
XX The peptide was used to demonstrate a novel method of protein/peptide
CC sequencing comprising sequentially degrading the peptide from the C
CC terminus by administering a vapour contg. perfluoroalkanoic acid and
CC analysing the resultant reaction mixt. The process is simple and avoids
CC using enzymes or other complicated organic reagents.
CC See also AAR32360-5.
CC
XX
SQ Sequence 8 AA;

Query Match 52.2%; Score 24; DB 14; Length 8;
Best Local Similarity 66.7%; Pred. No. 6.4e+05;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKLLIV 6
   1||:|
   3 philly 8

DB

Search completed: July 15, 2002, 13:38:45
Job time: 690 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:36:10 ; Search time 12.87 Seconds  
(without alignments)

15,183 Million cell updates/sec

Title: US-09-712-819A-4

Sequence: 1 PKLLIYMA 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 48605

Minimum DB seq length: 0  
Maximum DB seq length: 8

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCUTS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/Dockfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	58.5	8	3	US-07-961-307-1
2	22	47.8	5	3	US-08-981-122-11
3	22	47.8	6	1	US-08-180-209B-28
4	22	47.8	6	1	US-08-401-512-68
5	22	47.8	6	2	US-08-637-759B-295
6	22	47.8	6	3	US-08-871-355A-295
7	22	47.8	6	4	US-08-926-012-103
8	22	47.8	6	4	US-08-474-853-285
9	22	47.8	6	4	US-09-201-945-285
10	22	47.8	6	5	PCT-US94-02629-28
11	22	47.8	7	4	US-08-827-962-12
12	22	47.8	7	4	US-08-803-346-73
13	22	47.8	8	1	US-08-594-447-72
14	22	47.8	8	1	US-08-541-964-72
15	22	47.8	8	2	US-08-665-647-87
16	22	47.8	5	1	US-08-068-947-1
17	21	45.7	8	1	US-08-594-447-72
18	21	45.7	8	1	US-08-541-964-71
19	21	45.7	8	2	US-08-621-259A-205
20	21	45.7	8	2	US-08-665-647-86
21	21	45.7	8	2	US-08-031-538-32
22	21	45.7	8	6	516933-21
23	20	43.5	5	2	US-08-920-162A-29
24	20	43.5	5	4	US-09-356-931-29
25	20	43.5	5	5	PCT-US94-01321-48
26	20	43.5	6	4	US-09-128-572-4
27	20	43.5	6	6	5208144-9

28	20	43.5	6	6	5208144-10	Patent No. 5208144
29	20	43.5	7	1	US-08-036-210-1	Sequence 1, Appl
30	20	43.5	7	1	US-08-018-129-2	Sequence 2, Appl
31	20	43.5	7	1	US-08-018-129-12	Sequence 12, Appl
32	20	43.5	7	2	US-08-448-609-1	Sequence 1, Appl
33	20	43.5	7	2	US-08-448-250-2	Sequence 2, Appl
34	20	43.5	7	2	US-08-448-250-12	Sequence 12, Appl
35	20	43.5	7	2	US-08-836-480-13	Sequence 13, Appl
36	20	43.5	7	4	US-08-750-142B-50	Sequence 50, Appl
37	20	43.5	7	4	US-09-562-897-13	Sequence 13, Appl
38	20	43.5	8	5	PCT-US94-01321-46	Sequence 46, Appl
39	19	41.3	5	2	US-08-476-176B-50	Sequence 50, Appl
40	19	41.3	5	2	US-08-920-162A-18	Sequence 18, Appl
41	19	41.3	5	2	US-08-920-162A-32	Sequence 32, Appl
42	19	41.3	5	3	US-08-127-721A-50	Sequence 50, Appl
43	19	41.3	5	3	US-08-485-246A-50	Sequence 50, Appl
44	19	41.3	5	4	US-09-356-931-18	Sequence 18, Appl
45	19	41.3	5	4	US-09-356-931-32	Sequence 32, Appl

#### ALIGNMENTS

RESULT 1  
US-07-961-307-1  
Sequence 1, Application US/07961307  
Patent No. 6090785  
GENERAL INFORMATION:  
APPLICANT: Durelle, Philippe L.  
APPLICANT: Esser, Craig K.  
APPLICANT: Hagman, William K.  
APPLICANT: Kopka, Ihor E.  
TITLE OF INVENTION: Substituted N-Carboxyaryl-peptidals  
TITLE OF INVENTION: Derivatives as Antilegnerative Agents  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Merck & Co., Inc.  
STREET: 126 E. Lincoln Avenue, P. O. Box 2000  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/961,307  
FILING DATE: 15-OCT-1992  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Rose, David L.  
REGISTRATION NUMBER: 26332  
REFERENCE/DOCKET NUMBER: 16829  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908)594-4777  
TELEFAX: (908)594-4720  
TELEX: 138825  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
AMBI-SENSE: NO  
US-07-961-307-1  
Query Match 56.5%; Score 26; DB 3; Length 8;  
Best local Similarity 57.1%; Pred. No. 1.7e+05;

Matches 4: Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKLLIYW 7  
111:1  
DB 2 PKPLAFW 8

## RESULT 2

US-08-981-122-11  
Sequence 11, Application US/08981122B  
Patent No. 6127339  
GENERAL INFORMATION:  
APPLICANT: Hatanaka, Yoshihiro  
TITLE OF INVENTION: Peptide for binding cherezo a low density lipoprotein  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/08/981,122B  
CURRENT FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: JP 7-176904  
PRIOR FILING DATE: 1995-06-21  
PRIOR APPLICATION NUMBER: PCT/JP96/01734  
PRIOR FILING DATE: 1996-06-21  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 11  
LENGTH: 5  
TYPE: PPT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Sequence of a peptide synthesized in Example 2 from L-form  
Patent No. 6127339  
OTHER INFORMATION: F-moc amino acids by solid phase method using a multipetide  
US-08-981-122-11

## Query Match

Best Local Similarity 47.8%; Score 22; DB 3; Length 5;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 IYW 7  
111  
DB 3 IYW 5

## RESULT 3

US-08-180-209B-28  
Sequence 28, Application US/08180209B  
Patent No. 5593877  
GENERAL INFORMATION:  
APPLICANT: King, Te-Piao  
TITLE OF INVENTION: CLONING AND RECOMBINANT PRODUCTION OF  
TITLE OF INVENTION: VESPID VENOM ENZYMES, SUCH AS PHOSPHOLIPASE AND  
TITLE OF INVENTION: HYALURONIDASE, AND IMMUNOLOGICAL THERAPIES BASED  
TITLE OF INVENTION: THEREON  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/180,209B  
FILING DATE: 11-JAN-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/031,400

FILING DATE: 11-MAR-1993

ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-074 CIP

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: Internal

Query Match 47.8%; Score 22; DB 1; Length 6;

Best Local Similarity 100.0%; Pred. NO. 1.7e+05; Mismatches 0; Indels 0; Gaps 0;

QY 5 IYW 7  
111  
DB 3 IYW 5

## RESULT 4

US-08-401-512-68  
Sequence 68, Application US/08401512  
Patent No. 5596673  
GENERAL INFORMATION:  
APPLICANT: Keating, Mark T.  
TITLE OF INVENTION: Curran, Mark E.  
TITLE OF INVENTION: Wang, Qing  
TITLE OF INVENTION: Long QT Syndrome Genes  
NUMBER OF SEQUENCES: 81  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
STREET: 1201 New York Avenue, Suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3917  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/401,512  
FILING DATE: 09-MAR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Saxe, Stephen A.  
REGISTRATION NUMBER: 38,609  
REFERENCE/DOCKET NUMBER: 19780-113879  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-4848  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
US-08-401-512-68



Query Match 47.8%; Score 22; DB 1; Length 6;  
Best Local Similarity 40.0%; Pred. No. 1.7e+05;  
Matches 2; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 3 YMA 7  
| | | | |  
DB 1 LIVHW 5

## RESULT 5

US-08-637-759B-295  
; Sequence 295, Application US/08637759B  
; Patent No. 5876931  
; GENERAL INFORMATION:  
; APPLICANT: David William Holden  
; TITLE OF INVENTION: Identification of Genes  
; NUMBER OF SEQUENCES: 501  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/637,759B  
; FILING DATE: 03-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB95/02875  
; FILING DATE: 11-DEC-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; REFERENCE/DOCKET NUMBER: RPMS 101  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 873-8794  
; TELEFAX: (404) 873-8795  
; INFORMATION FOR SEQ ID NO: 295:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
US-08-637-759B-295

Query Match 47.8%; Score 22; DB 2; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 YMA 8  
| | | | |  
DB 3 YMA 5

## RESULT 6

US-08-871-355A-295  
; Sequence 295, Application US/08871355A  
; Patent No. 6015669  
; GENERAL INFORMATION:  
; APPLICANT: David William Holden

; TITLE OF INVENTION: Identification of Genes  
; NUMBER OF SEQUENCES: 501  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/871,355A  
; FILING DATE: 09-JUN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB95/02875  
; FILING DATE: 11-DEC-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; REFERENCE/DOCKET NUMBER: RPMS 101 CON  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 873-8794  
; TELEFAX: (404) 873-8795  
; INFORMATION FOR SEQ ID NO: 295:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
US-08-871-355A-295

## RESULT 7

Query Match 47.8%; Score 22; DB 3; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 YMA 8  
| | | | |  
DB 3 YMA 5

## RESULT 7

US-09-226-012-103  
; Sequence 103, Application US/09226012  
; Patent No. 6207383  
; GENERAL INFORMATION:  
; APPLICANT: Keating, Mark T.  
; TITLE OF INVENTION: MUTATIONS IN AND GENOMIC STRUCTURE OF HERG - A LONG QT  
; FILE REFERENCE: 2323-136  
; CURRENT APPLICATION NUMBER: US/09/226,012  
; EARLIER FILING DATE: 1999-01-06  
; CURRENT APPLICATION NUMBER: 09/122,847  
; NUMBER OF SEQ ID NOS: 116  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 103  
; LENGTH: 6  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-226-012-103

Query Match 47.8%; Score 22; DB 4; Length 6;  
Best Local Similarity 40.0%; Pred. No. 1.7e+05;  
Matches 2; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 3 IYIW 7  
1 IYIW 5

# RESULT 8

Sequence 28, Application US/08474853  
Patent No. 6287559  
GENERAL INFORMATION:  
APPLICANT: King, Te-piao  
TITLE OF INVENTION: CLONING AND RECOMBINANT PRODUCTION OF  
TITLE OF INVENTION: VESPID VENOM ENZYMES, SUCH AS PHOSPHOLIPASE AND  
TITLE OF INVENTION: HYALURONIDASE, AND IMMUNOLOGICAL THERAPIES BASED THEREON  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,853  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/180,209  
FILING DATE: 11-JAN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/031,400  
FILING DATE: 11-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION/DOCKET NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-074 CIPB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 343-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptidic  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
US-08-474-853-28

Query Match 47.8%; Score 22; DB 4; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 IYW 7  
1 IY 1  
Db 3 IYW 5

RESULT 9  
US-09-201-945-295

Sequence 295, Application US/09201945  
Patent No. 6342215  
GENERAL INFORMATION:  
APPLICANT: David William Holden  
TITLE OF INVENTION: Identification of Genes  
NUMBER OF SEQUENCES: 501  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/201,945  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/637,759  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION/DOCKET NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: RPN5 101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8795  
TELEFAX: (404) 873-8794  
INFORMATION FOR SEQ ID NO: 295:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-09-201-945-295

Query Match 47.8%; Score 22; DB 4; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 IYW 8  
1 IY 1  
Db 3 IYW 5

# RESULT 10

Sequence 28, Application PC/TUS9402629  
GENERAL INFORMATION:  
APPLICANT: King, Te-piao  
TITLE OF INVENTION: CLONING AND RECOMBINANT PRODUCTION OF  
TITLE OF INVENTION: VESPID VENOM ENZYMES, SUCH AS PHOSPHOLIPASE AND  
TITLE OF INVENTION: HYALURONIDASE, AND IMMUNOLOGICAL THERAPIES BASED THEREON  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/02629  
FILING DATE: 10-MAR-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/180,209  
FILING DATE: 11-JAN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/031,400  
FILING DATE: 11-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-074 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
AMTI-SENSE: NO  
FRAGMENT TYPE: internal  
PCT-US94-02629-28

Query Match 47.8%; Score 22; DB 5; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 IYW 7  
:||:  
Db 3 IYW 5

RESULT 11  
US-08-827-962-12  
Sequence 12, Application US/08827962A  
Patent No. 6258944  
GENERAL INFORMATION:  
APPLICANT: MERCK & CO., INC.  
TITLE OF INVENTION: OB RECEPTOR ISOFORMS AND NUCLEIC ACIDS  
FILE REFERENCE: 19693  
CURRENT APPLICATION NUMBER: US/08/827,962A  
CURRENT FILING DATE: 1997-05-06  
PRIOR APPLICATION NUMBER: 60/016,899  
PRIOR FILING DATE: 1996-05-06  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Mus Musculus  
US-08-827-962-12

Query Match 47.8%; Score 22; DB 4; Length 7;  
Best Local Similarity 50.0%; Pred. No. 1.7e+05;  
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 IYWA 8  
:||:  
Db 1 YWVS 4

RESULT 12  
US-08-803-346-73  
Sequence 73, Application US/08803346  
Patent No. 6281346  
GENERAL INFORMATION:  
APPLICANT: HESS, JOHN W.  
APPLICANT: CASKEY, C. THOMAS  
APPLICANT: LIU, QINGYUN  
APPLICANT: PHILLIPS, MICHAEL SEAN  
TITLE OF INVENTION: RAT OB RECEPTORS AND NUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 77  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JOANNE M. GIESSER - MERCK & CO., INC.  
STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000  
CITY: RAHWAY  
STATE: NJ  
COUNTRY: USA  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/803,346  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: GIESSER, JOANNE M  
REGISTRATION NUMBER: 32,838  
REFERENCE/DOCKET NUMBER: 19642Y  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 732-594-3046  
TELEFAX: 732-594-4720

INFORMATION FOR SEQ ID NO: 73:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-803-346-73

Query Match 47.8%; Score 22; DB 4; Length 7;  
Best Local Similarity 50.0%; Pred. No. 1.7e+05;  
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 IYWA 8  
:||:  
Db 1 YWVS 4

RESULT 13  
US-08-594-447-73  
Sequence 73, Application US/08594447  
Patent No. 5776716  
GENERAL INFORMATION:  
APPLICANT: Ron, Dorit  
APPLICANT: Napolitano, Eugene W.  
APPLICANT: Voronova, Anna F.  
TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH  
BLOCK THE INTERACTION OF FYN WITH PKC-THETA, AND USES  
TITLE OF INVENTION: THEROF  
NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FORSTER  
STREET: 2000 Pennsylvania Avenue, NW - Ste. 5500

CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/594,447  
FILING DATE: 31-JAN-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murashige, Kate H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22550-20025.24  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 822-0168  
TELEX: 90-4030 MRSNFORSMH  
INFORMATION FOR SEQ. ID NO: 73:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..8  
OTHER INFORMATION: /label= PRK2-2  
US-08-594-447-73

Query Match 47.88; Score 22; DB 1; Length 8;  
Best Local Similarity 33.38; Pred. No. 1.7e+05;  
Matches 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 KLLIYW 7  
: : : : :  
Db 1 EISVYW 6

RESULT 14  
US-08-541-964-72  
Sequence 72, Application US/08541964  
Patent No. 5783405  
GENERAL INFORMATION:  
APPLICANT: Mochly-Rosen, Daria  
APPLICANT: Ron, Dorit  
APPLICANT: Kaurav, Lawrence M.  
APPLICANT: Napolitano, Eugene W.  
TITLE OF INVENTION: A RAPID SCREENING METHOD FOR EFFECTORS  
NUMBER OF SEQUENCES: 74  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 PENNSYLVANIA AVENUE, NW-STE. 5500  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/541,964  
FILING DATE: 10-OCT-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Murashige, Kate H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22550-20025.23  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 822-0168  
TELEX: 90-4030 MRSNFORSMH  
INFORMATION FOR SEQ. ID NO: 72:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..8  
OTHER INFORMATION: /label= PRK2-2  
US-08-541-964-72

Query Match 47.88; Score 22; DB 1; Length 8;  
Best Local Similarity 33.38; Pred. No. 1.7e+05;  
Matches 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 KLLIYW 7  
: : : : :  
Db 1 EISVYW 6

RESULT 15  
US-08-665-647-87  
Sequence 87, Application US/08665647  
Patent No. 5935803  
GENERAL INFORMATION:  
APPLICANT: Pasquez, Nicki J.  
APPLICANT: Ron, Dorit  
APPLICANT: Voronova, Anna F.  
APPLICANT: Napolitano, Eugene W.  
TITLE OF INVENTION: METHODS TO IDENTIFY IMMUNOMODULATORS  
NUMBER OF SEQUENCES: 89  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 PENNSYLVANIA AVENUE, NW - Ste. 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/665,647  
FILING DATE: 18-JUN-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murashige, Kate H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22550-20025.25  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 822-0168  
TELEX: 90-4030 MRSNFORSMH  
INFORMATION FOR SEQ. ID NO: 87:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

```

; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..8
; OTHER INFORMATION: /label= PRK2-2
US-08-665-647-87

```

```

Query Match      47.84; Score 22; DB 2; Length 8;
Best Local Similarity 33.38; Pred. No. 17e+05;
Matches 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      2 KILTIW 7
       : : : :
Db      1 EISVY 6

```

Search completed: July 15, 2002, 13:39:25  
 Job time: 195 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:01:01 ; Search time 95.45 Seconds  
(without alignments)  
7.047 Million cell updates/sec

Title: US-09-712-819a-5

Perfect score: 35

Sequence: 1 TDFPLT 7

Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	100.0	34	2 H30607	Ig kappa chain V-I
2	35	100.0	54	2 A25521	Ig kappa chain V-I
3	35	100.0	54	2 J70521	Ig kappa chain V-I
4	35	100.0	71	2 S21526	Ig kappa chain V-I
5	35	100.0	78	2 S34102	Ig kappa chain V-I
6	35	100.0	79	2 S24215	Ig kappa chain V-I
7	35	100.0	81	2 PH1048	Ig light chain V-I
8	35	100.0	82	2 S38560	Ig light chain V-I
9	35	100.0	82	2 S34090	Ig kappa chain V-I
10	35	100.0	83	2 I30607	Ig kappa chain V-I
11	35	100.0	86	2 S16826	Ig kappa chain V-I
12	35	100.0	86	2 S16834	Ig kappa chain V-I
13	35	100.0	86	2 S16836	Ig kappa chain V-I
14	35	100.0	86	2 S34086	Ig kappa chain V-I
15	35	100.0	86	2 S16840	Ig kappa chain V-I
16	35	100.0	86	2 S16837	Ig kappa chain V-I
17	35	100.0	86	2 S16833	Ig kappa chain V-I
18	35	100.0	86	2 S16830	Ig kappa chain V-I
19	35	100.0	86	2 S16824	Ig kappa chain V-I
20	35	100.0	86	2 S16829	Ig kappa chain V-I
21	35	100.0	87	2 S16843	Ig kappa chain V-I
22	35	100.0	87	2 S16842	Ig kappa chain V-I
23	35	100.0	87	2 S16841	Ig kappa chain V-I
24	35	100.0	87	2 S34084	Ig kappa chain V-I
25	35	100.0	87	2 S34083	Ig kappa chain V-I
26	35	100.0	88	2 S21528	Ig kappa chain V-I
27	35	100.0	88	2 S21524	Ig kappa chain V-I
28	35	100.0	88	2 S21525	Ig kappa chain V-I
29	35	100.0	88	2 S21522	Ig kappa chain V-I

30	35	100.0	89	2 S34096	Ig kappa chain V-I
31	35	100.0	90	2 S38561	Ig light chain V-I
32	35	100.0	90	2 I38601	Ig kappa chain V-I
33	35	100.0	91	2 PH1071	Ig light chain V-I
34	35	100.0	91	2 S25462	Ig kappa chain V-I
35	35	100.0	91	2 S37520	Ig kappa chain V-I
36	35	100.0	91	2 S37521	Ig kappa chain V-I
37	35	100.0	91	2 S67940	Ig kappa chain V-I
38	35	100.0	92	2 S37533	Ig kappa chain V-I
39	35	100.0	92	2 S37530	Ig kappa chain V-I
40	35	100.0	92	2 S37529	Ig kappa chain V-I
41	35	100.0	92	2 S37535	Ig kappa chain V-I
42	35	100.0	92	2 S37524	Ig kappa chain V-I
43	35	100.0	92	2 S37513	Ig kappa chain V-I
44	35	100.0	92	2 S37512	Ig kappa chain V-I
45	35	100.0	92	2 S37531	Ig kappa chain V-I

#### ALIGNMENTS

RESULT 1  
H30607  
Ig kappa chain V-I region (Bla) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 29-Jun-1989 #sequence\_revision 29-Jun-1989 #text\_change 30-May-1997  
C:Accession: H30607  
R:Gonl, F.R.; Chen, P.P.; McGinns, D.; Arjouni, M.L.; Fernandez, J.; Carson, D.; S  
J. Immunol. 142, 3158-3163, 1989  
A:Title: Structural and idiotype characterization of the L chains of human Igm autoa  
A:Reference number: A30601; MID:89215279  
A:Accession: H30607  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-34 <GON>  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 35; DB 2; Length 34;  
Best Local Similarity 100.0%; Pred. No. 0.44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFPLT 7  
DB 23 TDFPLT 29

RESULT 2  
A25521  
Ig kappa chain V region (321) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 21-May-1988 #sequence\_revision 21-May-1990 #text\_change 09-May-1997  
C:Accession: A25521  
R:Chen, P.P.; Albrandt, K.; Orida, M.K.; Padoux, V.; Chen, E.Y.; Schreuter, R.; Liu, J  
Proc. Natl. Acad. Sci. U.S.A. 83, 8318-8322, 1986  
A:Title: Genetic basis for the cross-reactive Idiotypes on the light chains of human  
A:Reference number: A94135; MID:87041448  
A:Accession: A25521  
A:Molecule type: DNA  
A:Residues: 1-54 <CHB>  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 35; DB 2; Length 54;  
Best Local Similarity 100.0%; Pred. No. 0.71;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFPLT 7  
DB 28 TDFPLT 34

## RESULT 3

JT0521

Ig kappa chain V-III region (CPI) - human (fragment)

C:Species: Homo sapiens (man)

C&gt;Date: 23-Oct-1992 #sequence\_revision 23-Oct-1992 #text\_change 09-May-1997

C:Accession: JT0521

R:Anker, R.; Conley, M.E.; Pollok, B.A.

J. Exp. Med. 169, 2109-2119, 1989

A&gt;Title: Clonal diversity in the B cell repertoire of patients with X-linked agammaglobulinemia

A:Reference number: JT0521; MUID:89279157

A:Accession: JT0521

A:Molecule type: mRNA

A:Residues: 1-54 &lt;ANK&gt;

A&gt;Note: The sequence shown here is one of eight productive V-D-J mu chain rearrangements

C:Superfamily: Immunoglobulin V region; Immunoglobulin homology

C:Keywords: heterotetramer; Immunoglobulin

F:1-49/Domain: V region &lt;VRE&gt;

F:50-54/Domain: J region &lt;JRE&gt;

## Query Match

100.0%; Score 35; DB 2; Length 54;

Best Local Similarity 100.0%; Pred. No. 0.71;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7

|||||

Db 21 TDFTLTI 27

## RESULT 4

S21526

Ig kappa chain V region - human (fragment)

C:Species: Homo sapiens (man)

C&gt;Date: 20-Feb-1995 #sequence\_revision 25-Oct-1996 #text\_change 23-Jul-1999

C:Accession: S34082; S21526

R:Magner, S.D.; Luzzatto, L.

Eur. J. Immunol. 23, 391-397, 1993

A&gt;Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distributed

A:Reference number: S34076; MUID:93170387

A:Accession: S34082

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-71 &lt;MAZ&gt;

A:Cross-references: EMBL:X66042; NID:933318; PIDN:CAA46841.1; PID:933319

C:Superfamily: Immunoglobulin V region; Immunoglobulin homology

C:Keywords: heterotetramer; Immunoglobulin

## Query Match

100.0%; Score 35; DB 2; Length 71;

Best Local Similarity 100.0%; Pred. No. 0.95;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7

|||||

Db 62 TDFTLTI 68

## RESULT 5

S34102

Ig kappa chain V region - human

C:Species: Homo sapiens (man)

C&gt;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 07-May-1999

C:Accession: S34102

R:Magner, S.D.; Luzzatto, L.

Eur. J. Immunol. 23, 391-397, 1993

A&gt;Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distributed

A:Reference number: S34076; MUID:93170387

A:Accession: S34102

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-78 &lt;WAG&gt;

A:Cross-references: EMBL:X67186

C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin

## Query Match

100.0%; Score 35; DB 2; Length 78;

Best Local Similarity 100.0%; Pred. No. 1;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7

|||||

Db 60 TDFTLTI 66

## RESULT 6

S24215

Ig kappa chain - mouse (fragment)

C:Species: Mus musculus (house mouse)

C&gt;Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999

C:Accession: S24215

R:Shimizu, T.; Iwasato, T.; Yamagishi, H.

J. Exp. Med. 173, 1065-1072, 1991

A&gt;Title: Deletions of immunoglobulin C(kappa) region characterized by the circular ex

A:Reference number: S24214; MUID:91217618

A:Accession: S24215

A:Status: preliminary; translation not shown

A:Molecule type: DNA

A:Residues: 1-79 &lt;SHI&gt;

A:Cross-references: EMBL:X58202; NID:953718; PIDN:CAA41178.1; PID:9930195

C:Superfamily: Immunoglobulin V region; Immunoglobulin homology

C:Keywords: heterotetramer; Immunoglobulin

## Query Match

100.0%; Score 35; DB 2; Length 79;

Best Local Similarity 100.0%; Pred. No. 1.1;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7

|||||

Db 42 TDFTLTI 48

## RESULT 7

PH1048

Ig light chain V region (clone 165.49) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C&gt;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 15-Jun-1996

C:Accession: PH1048

R:Tillman, D.M.; Jou, N.T.; Hiji, R.J.; Marlon, T.N.

J. Exp. Med. 176, 761-779, 1992

A&gt;Title: Both IgM and IgG anti-DNA antibodies are the products of clonally selective

A:Reference number: PH0971; MUID:92381444

A:Accession: PH1048

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-81 &lt;TIL&gt;

A:Experimental source: B cell, strain [N2B x NZW]F1

C:Superfamily: Immunoglobulin V region; Immunoglobulin homology

C:Keywords: immunoglobulin

## Query Match

100.0%; Score 35; DB 2; Length 81;

Best Local Similarity 100.0%; Pred. No. 1.1;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7

|||||

Db 54 TDFTLTI 60

## RESULT 8

S38560

Ig light chain V region (ASMB1) - mouse (fragment)

C:Species: Mus musculus (house mouse)



C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999  
 C:Accession: S38560  
 R:Monestier, M.; Lozman, L.J.; Novick, K.E.; Aris, J.P.  
 submitted to the EMBL Data Library, September 1993  
 A:Description: Molecular analysis of mercury-induced anti-nucleolar antibodies in H-2s M  
 A:Reference number: S38559  
 A:Accession: S38560  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-82 <MON>  
 A:Cross-references: EMBL:X75102; NID:9414147; PIDN:CAAS2993.1; PID:9414148  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: immunoglobulin

Query Match 100.0%; Score 35; DB 2; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
 |||||  
 DB 69 TDFTLTI 75

RESULT 9  
 S34090  
 Ig kappa chain V region - human  
 C:Species: Homo sapiens (man)  
 C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 21-Jan-2000  
 C:Accession: S34090  
 R:Magner, S.D.; Luzzatto, L.  
 Eur. J. Immunol. 23, 391-397, 1993  
 A:Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distribute  
 A:Reference number: S34076; MUID:93170387  
 A:Accession: S34090  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-82 <WAG>  
 A:Cross-references: EMBL:X67174  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:7-81/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 35; DB 2; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
 |||||  
 DB 60 TDFTLTI 66

RESULT 10  
 I30607  
 Ig kappa chain V-III region (We1) - human (fragments)  
 C:Species: Homo sapiens (man)  
 C>Date: 29-Jun-1989 #sequence\_revision 29-Jun-1989 #text\_change 09-May-1997  
 C:Accession: I30607  
 R:Goni, F.R.; Chen, P.P.; McGinnis, D.; Arionilla, M.L.; Fernandez, J.; Carson, D.; Soli  
 J. Immunol. 142, 3158-3163, 1989  
 A:Title: Structural and idiotypic characterization of the L chains of human Igm autoanti  
 A:Reference number: A30601; MUID:89215279  
 A:Accession: I30607  
 A>Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-83 <CON>  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 35; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 1.1;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 TDFTLTI 7  
 |||||  
 DB 70 TDFTLTI 76

RESULT 11  
 S16826  
 Ig kappa chain V region - human (fragment)  
 C:Species: Homo sapiens (man)  
 C>Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000  
 C:Accession: S16826; S34101  
 R:Blaiss, G.; Kuntz, J.L.; Pasquali, J.L.  
 Eur. J. Immunol. 21, 1221-1227, 1991  
 A:Title: Molecular analysis of V(kappa)III variable regions of polyclonal rheumatoid  
 A:Reference number: S16823; MUID:91243737  
 A:Accession: S16826  
 A>Status: preliminary; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-86 <BLA>  
 A:Cross-references: EMBL:X54824; NID:933653; PIDN:CAAS38593.1; PID:933654  
 R:Magner, S.D.; Luzzatto, L.  
 Eur. J. Immunol. 23, 391-397, 1993  
 A:Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distrib  
 A:Reference number: S34076; MUID:93170387  
 A:Accession: S34101  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-86 <WAG>  
 A:Cross-references: EMBL:X67185  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:7-81/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 35; DB 2; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
 |||||  
 DB 60 TDFTLTI 66

RESULT 12  
 S16834  
 Ig kappa chain V region - human (fragment)  
 C:Species: Homo sapiens (man)  
 C>Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000  
 C:Accession: S16834  
 R:Blaiss, G.; Kuntz, J.L.; Pasquali, J.L.  
 Eur. J. Immunol. 21, 1221-1227, 1991  
 A:Title: Molecular analysis of V(kappa)III variable regions of polyclonal rheumatoid  
 A:Reference number: S16823; MUID:91243737  
 A:Accession: S16834  
 A>Status: preliminary; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-86 <BLA>  
 A:Cross-references: EMBL:X54832  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:7-81/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 35; DB 2; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
 |||||  
 DB 60 TDFTLTI 66

RESULT 13  
S16836  
Ig kappa chain V region - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000  
C:Accession: S16836  
R:Blaisson, G.; Kuntz, J.L.; Pasquall, J.L.  
Eur. J. Immunol. 21, 1221-1227, 1991  
A:Title: Molecular analysis of V(kappa)III variable regions of polyclonal rheumatoid fac  
A:Reference number: S16823; MUID:91243737  
A:Accession: S16836  
A:Status: preliminary; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-86 <BLA>  
A:Cross-references: EMBL:X54834  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:7-81/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 35; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
|||||  
DB 60 TDFTLTI 66

RESULT 14  
S34086  
Ig kappa chain V region - human  
C:Species: Homo sapiens (man)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 21-Jan-2000  
C:Accession: S34086  
R:Wagner, S.D.; Luzzatto, L.  
Eur. J. Immunol. 23, 391-397, 1993  
A:Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distributed  
A:Reference number: S34076; MUID:93170387  
A:Accession: S34086  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-86 <NAG>  
A:Cross-references: EMBL:X67169  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:9-83/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 35; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
|||||  
DB 62 TDFTLTI 68

RESULT 15  
S16840  
Ig kappa chain V region - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000  
C:Accession: S16840  
R:Blaisson, G.; Kuntz, J.L.; Pasquall, J.L.  
Eur. J. Immunol. 21, 1221-1227, 1991  
A:Title: Molecular analysis of V(kappa)III variable regions of polyclonal rheumatoid fac  
A:Reference number: S16823; MUID:91243737  
A:Accession: S16840  
A:Status: preliminary; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-86 <BLA>

A:Cross-references: EMBL:X54838  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:7-81/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 35; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
|||||  
DB 60 TDFTLTI 66

Search completed: July 15, 2002, 13:01:01  
Job time: 467 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:23:32 ; Search time 44.9 Seconds  
(without alignments)  
6.036 Million cell updates/sec

Title: us-09-712-819a-5

Perfect score: 35  
Sequence: 1 TDFTLTI 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	35	100.0	100	1 KV3C_HUMAN	P01621 homo sapien
2	35	100.0	107	1 KV04_RABIT	P01685 oryctolagus
3	35	100.0	107	1 KV1D_HUMAN	P01596 homo sapien
4	35	100.0	108	1 KV06_RABIT	P01687 oryctolagus
5	35	100.0	108	1 KV1E_HUMAN	P01597 homo sapien
6	35	100.0	108	1 KV1H_HUMAN	P01600 homo sapien
7	35	100.0	108	1 KV1O_HUMAN	P01609 homo sapien
8	35	100.0	108	1 KV5P_MOUSE	P01649 mus musculu
9	35	100.0	108	1 KV5R_MOUSE	P01650 mus musculu
10	35	100.0	108	1 KV5S_MOUSE	P01651 mus musculu
11	35	100.0	108	1 KV5T_MOUSE	P01652 mus musculu
12	35	100.0	108	1 KV5T_MOUSE	P01653 mus musculu
13	35	100.0	109	1 KV01_RAT	P01681 rattus norv
14	35	100.0	109	1 KV1T_HUMAN	P01612 homo sapien
15	35	100.0	109	1 KV3B_HUMAN	P01620 homo sapien
16	35	100.0	109	1 KV3D_HUMAN	P01622 homo sapien
17	35	100.0	109	1 KV3E_HUMAN	P01623 homo sapien
18	35	100.0	109	1 KV3G_HUMAN	P04206 homo sapien
19	35	100.0	111	1 KV3H_MOUSE	P01660 mus musculu
20	35	100.0	111	1 KV3J_MOUSE	P01662 mus musculu
21	35	100.0	111	1 KV3K_MOUSE	P01663 mus musculu
22	35	100.0	114	1 KV1A_MOUSE	P01632 mus musculu
23	35	100.0	114	1 KV4A_HUMAN	P01625 homo sapien
24	35	100.0	115	1 KV3I_HUMAN	P04433 homo sapien
25	35	100.0	115	1 KV5A_HUMAN	P06315 homo sapien
26	35	100.0	116	1 KV3J_HUMAN	P04434 homo sapien
27	35	100.0	117	1 KV4I_HUMAN	P01601 homo sapien
28	35	100.0	121	1 KV4O_HUMAN	P06315 homo sapien
29	35	100.0	129	1 KV1K_HUMAN	P04431 homo sapien
30	35	100.0	129	1 KV3L_HUMAN	P01815 homo sapien
31	35	100.0	129	1 KV3M_HUMAN	P01816 homo sapien
32	35	100.0	131	1 KV4J_MOUSE	P01661 mus musculu
33	35	100.0	133	1 KV4B_HUMAN	P06313 homo sapien

34	35	100.0	134	1 KV4C_HUMAN	P06314 homo sapien
35	35	100.0	136	1 KV5B_MOUSE	P01634 mus musculu
36	33	94.3	104	1 KV17_RABIT	P01688 oryctolagus
37	33	94.3	112	1 KV3G_MOUSE	P01659 mus musculu
38	33	94.3	132	1 KV3F_MOUSE	P01658 mus musculu
39	32	91.4	285	1 ALXM_PHOS4	P39049 photobacter
40	31	88.6	108	1 KV05_RABIT	P01686 oryctolagus
41	31	88.6	108	1 KV07_RABIT	P01689 oryctolagus
42	31	88.6	108	1 KV08_RABIT	P01689 oryctolagus
43	31	88.6	108	1 KV1A_HUMAN	P01593 homo sapien
44	31	88.6	108	1 KV1E_HUMAN	P01598 homo sapien
45	31	88.6	108	1 KV1G_HUMAN	P01599 homo sapien

## ALIGNMENTS

```

RESULT 1
KV3C_HUMAN          STANDARD:      PRT:   100 AA.
ID  KV3C_HUMAN
AC  P01621;
DT  21-JUL-1986 (Rel. 01, Created)
DE  21-JUL-1986 (Rel. 01, Last sequence update)
DI  15-JUL-1999 (Rel. 38, Last annotation update)
DE  Ig kappa chain V-TII region NG9 precursor (Fragment).
OS  Homo sapiens (Human)
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=84093600; PubMed=6419127;
RA  Bentley D.L.;
RT  "Best kappa immunoglobulin mRNA in human lymphocytes is homologous to
   a small family of germ-line V genes.";
RL  Nature 307:77-80(1984).
CC  -i- MISCELLANEOUS: THIS GENE WAS ISOLATED FROM THE NG9/9.1 HYBRIDOMA.
DR  PIR: A01894; K3HONG.
DR  HSSP: P01789; IMCP.
DR  InterPro: IPR003006; Ig_MHC.
DR  InterPro: IPR003596; Ig_V.
DR  Pfam: PF00047; Ig; 1.
DR  SMART: SM00406; Ig; 1.
KW  Immunoglobulin V region; Signal; Hydridoma.
FT  NON_TER
FT  SIGNAL
FT  CHAIN
FT  DISULFD
FT  NON_TER
SQ  SEQUENCE 100 AA; 5D9AF363C52632F CRC64;

Query Match          100.0%; Score 35; DB 1; Length 100;
Best local similarity 100.0%; Pred. No. 0.32; Mismatches 0; Gaps 0;
Matches 7; Conservative 0; Indels 0;

QY  1 TDFTLTI 7
Db  74 TDFTLTI 80

RESULT 2
KV04_RABIT          STANDARD:      PRT:   107 AA.
ID  KV04_RABIT
AC  P01685;
DT  21-JUL-1986 (Rel. 01, Created)
DT  21-JUL-1986 (Rel. 01, Last sequence update)
DI  15-JUL-1999 (Rel. 38, Last annotation update)
DE  Ig kappa chain V region 4135.
OS  Oryctolagus cuniculus (Rabbit).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX  NCBI_TaxID=9986;

```

RN [1]  
 RP MEDLINE=75133568; PubMed=1091650;  
 RX Best Local Similarity 100.0%; Pred. No. 0.34;  
 RA "Primary structure of the L chain from a rabbit homogeneous antibody  
 RT to streptococcal carbohydrate. II. Sequence determination of peptides  
 RT from tryptic and peptic digests."  
 RL J. Biol. Chem. 250:3289-3296(1975).  
 CC -1- MISCELLANEOUS: THE SEQUENCE OF THE B4-TYPE C REGION IS ALSO GIVEN.  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM ANTIBODY TO THE  
 CC SPECIFIC CARBOHYDRATE OF GROUP C STREPTOCOCCI AND WAS ISOLATED  
 CC FROM THE SERUM OF A SINGLE RABBIT.  
 DR PIR: A01948; K4RB41.  
 DR HSSP: P80362; 1WTU.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_V.  
 DR Pfam: PF00047; Ig\_1.  
 DR SMART: SM00406; IgV\_1.  
 KW Immunoglobulin V region.  
 FT DOMAIN 1 24 FRAMEWORK-1.  
 FT DOMAIN 25 35 COMPLEMENTARITY-DETERMINING-1.  
 FT DOMAIN 36 50 FRAMEWORK-2.  
 FT DOMAIN 51 57 COMPLEMENTARITY-DETERMINING-2.  
 FT DOMAIN 58 89 FRAMEWORK-3.  
 FT DOMAIN 90 96 COMPLEMENTARITY-DETERMINING-3.  
 FT DOMAIN 97 106 FRAMEWORK-4.  
 FT NON\_TER 107 107  
 SQ SEQUENCE 107 AA; 11182 MW; 8F84C5FEE60B7222 CRC64;

Query Match 100.0%; Score 35; DB 1; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 0.34;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFTLTI 7  
 DB 70 TDFTLTI 76  
 RESULT 3  
 KYLE\_HUMAN STANDARD; PRT; 107 AA.  
 AC P01596;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain V-I region CAR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=75075135; PubMed=4216454;  
 RA Mistein C.P., Deverson E.V.;  
 RT "Primary structure of kappa light chain from a human myeloma  
 RT protein."  
 RL Eur. J. Biochem. 49:377-391(1974).  
 CC -1- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (1,2)  
 CC MARKER.  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN.  
 DR PIR: A01864; K1HUAR.  
 DR HSSP: P80362; 1WTU.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_V.  
 DR Pfam: PF00047; Ig\_1.  
 DR SMART: SM00406; IgV\_1.  
 KW Immunoglobulin V region; Glycoprotein.  
 KM CARBOHYD 28 28 N-LINKED (GLCNAC. . .).  
 FT NON\_TER 107 107  
 SQ SEQUENCE 107 AA; 11703 MW; E1BF0DF9844C346 CRC64;

Query Match 100.0%; Score 35; DB 1; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 0.34;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFTLTI 7  
 DB 69 TDFTLTI 75  
 RESULT 4  
 KY06\_RABIT STANDARD; PRT; 108 AA.  
 ID P01687;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain V region BS-5.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_Taxid=9986;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=75127348; PubMed=4156171;  
 RA Jaton J.-C.;  
 RT "Comparison of the amino acid sequences of the variable regions of  
 RT light chains derived from two homogeneous rabbit anti-pneumococcal  
 RT antibodies."  
 RL Biochem. J. 141:15-25(1974).  
 RN [2]  
 RP AMIDES AT POSITIONS 37 AND 89.  
 RA Jaton J.-C.;  
 RL Submitted (JUN-1975) to the PIR data bank.  
 CC -1- MISCELLANEOUS: THIS CHAIN DIFFERS FROM THE KAPPA CHAIN FROM RABBIT  
 CC BS-1, ALSO OBTAINED FROM ANTIBODY TO TYPE III PNEUMOCOCCI, AT 8  
 CC POSITIONS IN THE V REGION.  
 DR PIR: A01950; KYRBS5.  
 DR HSSP: P01789; 2MCP.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_V.  
 DR Pfam: PF00047; Ig\_1.  
 DR SMART: SM00406; IgV\_1.  
 KW Immunoglobulin V region.  
 FT DOMAIN 1 23 FRAMEWORK-1.  
 FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.  
 FT DOMAIN 35 49 FRAMEWORK-2.  
 FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.  
 FT DOMAIN 57 88 FRAMEWORK-3.  
 FT DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-3.  
 FT DOMAIN 98 107 FRAMEWORK-4.  
 FT DISULFID 23 88 WITH A CYS IN THE C REGION.  
 FT NON\_TER 108 108  
 SQ SEQUENCE 108 AA; 11281 MW; CFB6D10DBB7A5FCE CRC64;

Query Match 100.0%; Score 35; DB 1; Length 108;  
 Best Local Similarity 100.0%; Pred. No. 0.35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFTLTI 7  
 DB 69 TDFTLTI 75  
 RESULT 5  
 KYLE\_HUMAN STANDARD; PRT; 108 AA.  
 AC P01597;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain V-I region DEE.

```

OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE
RX MEDLINE=72053133; PubMed=5124396;
RA Milstein C.P.; Deverson E.V.;
RT "The amino acid sequence of a human kappa light chain.";
RL Biochem. J. 123:945-958(1971).
CC -1- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (3) MARKER.
DR PIR: A01865; KIHUDE.
DR HSSP: P01607; 1REI.
DR InterPro: IPR003006; I9_MHC.
DR InterPro: IPR003596; I9_V.
DR Pfam: PF00047; I9; 1.
DR SMART: SM00406; Igv; 1.
KM Immunoglobulin V region.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 35 49 FRAMEWORK-2.
FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 57 88 FRAMEWORK-3.
FT DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 98 107 FRAMEWORK-4.
FT DISULFID 23 88 BY SIMILARITY.
FT NON_TER 108 108
SQ SEQUENCE 108 AA; 11661 MW; BDD6E350017FE1E51 CRC64;

```

```

Query Match 100.0%; Score 35; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 0.35; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;
QY 1 TDEFTLTI 7
DB 69 TDEFTLTI 75

```

```

RESULT 6
ID KVIH_HUMAN STANDARD; PRT; 108 AA.
AC P01600;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-I region Hau.
OS Homo sapiens (human); Chordata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=71032830; PubMed=4097974;
RA Watanabe S.; Hilschmann N.;
RT "The primary structure of a monoclonal kappa-type immunoglobulin L-
chain of subgroup I (Bence-Jones protein Hau): subdivision within
subgroups.";
RT Hoppe-Seyler's Z. Physiol. Chem. 351:1291-1295(1970).
CC -1- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (3) MARKER.
CC -1- MISCELLANEOUS: THIS IS A BENCE-JONES PROTEIN.
DR PIR: A01868; KIHOUH.
DR HSSP: P80362; 1MTL.
DR InterPro: IPR003006; I9_MHC.
DR InterPro: IPR003596; I9_V.
DR Pfam: PF00047; I9; 1.
DR SMART: SM00406; Igv; 1.
KM Immunoglobulin V region; Bence-Jones protein.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 35 49 FRAMEWORK-2.
FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 57 88 FRAMEWORK-3.

```

```

FT DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 98 107 FRAMEWORK-4.
FT DISULFID 23 88 BY SIMILARITY.
FT NON_TER 108 108
SQ SEQUENCE 108 AA; 11671 MW; 08D3A61608D0618 CRC64;

```

```

Query Match 100.0%; Score 35; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 0.35; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;
QY 1 TDEFTLTI 7
DB 69 TDEFTLTI 75

```

```

RESULT 7
ID KVIQ_HUMAN STANDARD; PRT; 108 AA.
AC P01609;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-I region Scw.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=75059271; PubMed=4435756;
RA Eulitz M.; Hilschmann N.;
RT "The primary structure of a human immunoglobulin L-chain of
kappa-type (Bence-Jones protein Scw); II: The chymotryptic peptides
RT and the complete amino acid sequence.";
RT Hoppe-Seyler's Z. Physiol. Chem. 355:842-866(1974).
CC -1- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (1,2)
MARKER.
CC -1- MISCELLANEOUS: THIS IS A BENCE-JONES PROTEIN.
DR PIR: A01875; KIHUSW.
DR HSSP: P01607; 1REI.
DR InterPro: IPR003006; I9_MHC.
DR InterPro: IPR003596; I9_V.
DR Pfam: PF00047; I9; 1.
DR SMART: SM00406; Igv; 1.
KM Immunoglobulin V region; Bence-Jones protein.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 35 49 FRAMEWORK-2.
FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 57 88 FRAMEWORK-3.
FT DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 98 107 FRAMEWORK-4.
FT DISULFID 23 88 BY SIMILARITY.
FT NON_TER 108 108
SQ SEQUENCE 108 AA; 11764 MW; 32CECDDE9644414 CRC64;

```

```

Query Match 100.0%; Score 35; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 0.35; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;
QY 1 TDEFTLTI 7
DB 69 TDEFTLTI 75

```

```

RESULT 8
ID KVSF_MOUSE STANDARD; PRT; 108 AA.
AC F01649;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)

```

```

DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Ig kappa chain V-V regions (Anti-arsenate antibodies).
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE.
RC STRAIN-A/J;
RX MEDLINE=71250895; PubMed=70482;
RA Capra J.D., Tung A.S., Nisenooff A.;
RT Structural studies on induced antibodies with defined idiotypic
RT specificities. V. The complete amino acid sequence of the light chain
RT variable regions of anti-p-azophenylarsenate antibodies from A/J mice
RT bearing a cross-reactive idiotype."
RL J. Immunol. 119:993-999(1977).
CC -1- MISCELLANEOUS: THE MIXTURE SEQUENCED CONTAINED AT LEAST TWO OR
CC THREE DIFFERENT LIGHT CHAINS. PEPTIDES CONTAINING THE FOLLOWING
CC SUBSTITUTIONS WERE ALSO ISOLATED: 3-VAL, 10-ILE, AND 12-LEU;
CC 22-LEU; 36-PHE, 41-GLU, AND 43-ALA; 63-THR AND 68-ARG; 76-SER,
CC 77-SER, 78-VAL, 80-ALA, AND 85-ASP; AND 100-GLN AND 107-ARG.
DR PIR: A01928; KMSA.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00406; IgV_1.
KW Immunoglobulin V region; Antiarsonate antibody.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 35 49 FRAMEWORK-2.
FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 57 88 FRAMEWORK-3.
FT DOMAIN 89 97 FRAMEWORK-4.
FT DOMAIN 98 108 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 23 88 BY SIMILARITY.
FT NON_TER 108
SQ SEQUENCE 108 AA; 12056 MW; AE2861E6AAC09DD2 CRC64;

Query Match 100.0%; Score 35; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFTLTI 7
|111111
DB 69 TDFTLTI 75

RESULT 9
KV5Q_MOUSE STANDARD; PRT; 108 AA.
ID KV5Q_MOUSE
AC P01650;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-V region UPC 61.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=79195288; PubMed=109517;
RA Viana M., Rudikoff S., Potter M.;
RT The structural basis of a hapten-inhibitable kappa-chain idiotype."
RL J. Immunol. 122:1905-1910(1979).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT
CC BIND BETA(2-1)-FRUCTOFURANOSYL MOIETIES (INULIN).
DR PIR: A01929; KMS61.
DR HSR: P00362; IRTL.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_1.

```

```

DR SMART: SM00406; IgV_1.
KW Immunoglobulin V region.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 35 49 FRAMEWORK-2.
FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 57 88 FRAMEWORK-3.
FT DOMAIN 89 97 FRAMEWORK-4.
FT DOMAIN 98 108 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 23 88 BY SIMILARITY.
FT NON_TER 108
SQ SEQUENCE 108 AA; 11809 MW; FA4DA36076F2AFE CRC64;

Query Match 100.0%; Score 35; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFTLTI 7
|111111
DB 69 TDFTLTI 75

RESULT 10
KV5R_MOUSE STANDARD; PRT; 108 AA.
ID KV5R_MOUSE
AC P01651;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-V region EPC 109.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=79195288; PubMed=109517;
RA Viana M., Rudikoff S., Potter M.;
RT The structural basis of a hapten-inhibitable kappa-chain idiotype."
RL J. Immunol. 122:1905-1910(1979).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT
CC BIND BETA(2-1)-FRUCTOFURANOSYL MOIETIES (INULIN).
DR PIR: B92808; KMS09.
DR HSR: P01607; IRTI.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00406; IgV_1.
KW Immunoglobulin V region.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 35 49 FRAMEWORK-2.
FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 57 88 FRAMEWORK-3.
FT DOMAIN 89 97 FRAMEWORK-4.
FT DOMAIN 98 108 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 23 88 BY SIMILARITY.
FT NON_TER 108
SQ SEQUENCE 108 AA; 11876 MW; 35C116BD60F79310 CRC64;

Query Match 100.0%; Score 35; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFTLTI 7
|111111
DB 69 TDFTLTI 75

RESULT 11
KV5S_MOUSE

```

```

ID  KY5_MOUSE      STANDARD:      PRT:  108 AA.
AC  P01652;
DT  21-JUL-1986 (Rel. 01, Created)
DT  21-JUL-1986 (Rel. 01, Last sequence update)
DT  15-JUL-1999 (Rel. 38, Last annotation update)
DE  Ig kappa chain V-V region J606.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE.
RA  MEDLINE=82099361; PubMed=6798111;
RA  Johnson N., Slankard J., Paul L., Hood L.;
RT  "The complete V domain amino acid sequences of two myeloma Inulin-
RT  binding proteins."
RL  J. Immunol. 128:302-307(1982).
CC  -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT
CC  BIND BETA(2-1)-FRUCTOFURANOSYL MOIETIES (INULIN).
DR  PIR; A92811; KWS06.
DR  HSSP; P01607; IREI.
DR  InterPro; IPR003006; Ig_MHC.
DR  InterPro; IPR003596; Ig_V.
DR  Pfam; PF00047; Ig; 1.
DR  SMART; SM00406; IGV; 1.
KM  Immunoglobulin V region.
FT  DOMAIN 1 23 FRAMEWORK-1.
FT  DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.
FT  DOMAIN 35 49 FRAMEWORK-2.
FT  DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.
FT  DOMAIN 57 88 FRAMEWORK-3.
FT  DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-3.
FT  DOMAIN 98 108 FRAMEWORK-4.
FT  DISUFID 23 88 BY SIMILARITY.
FT  NON_TER 108
SQ  SEQUENCE 108 AA; 11810 MW; 8DE4DD31076F2AFB CRC64;

Query Match 100.0%; Score 35; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 TDFTLTI 7
    |||||
DB  69 TDFTLTI 75

RESULT 12
KVIT_MOUSE
ID  KVIT_MOUSE      STANDARD:      PRT:  108 AA.
AC  P01653;
DT  21-JUL-1986 (Rel. 01, Created)
DT  21-JUL-1986 (Rel. 01, Last sequence update)
DT  15-JUL-1999 (Rel. 38, Last annotation update)
DE  Ig kappa chain V-V region W3082.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE.
RA  MEDLINE=82099361; PubMed=6798111;
RA  Johnson N., Slankard J., Paul L., Hood L.;
RT  "The complete V domain amino acid sequences of two myeloma Inulin-
RT  binding proteins."
RL  J. Immunol. 128:302-307(1982).
CC  -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT
CC  BIND BETA(2-1)-FRUCTOFURANOSYL MOIETIES (INULIN).
DR  PIR; B92811; KWS82.
DR  HSSP; P80362; IWTI.
DR  InterPro; IPR003006; Ig_MHC.
DR  InterPro; IPR003596; Ig_V.
DR  Pfam; PF00047; Ig; 1.

```

```

DR  SMART; SM00406; IGV; 1.
KM  Immunoglobulin V region.
FT  DOMAIN 1 23 FRAMEWORK-1.
FT  DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.
FT  DOMAIN 35 49 FRAMEWORK-2.
FT  DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.
FT  DOMAIN 57 88 FRAMEWORK-3.
FT  DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-3.
FT  DOMAIN 98 108 FRAMEWORK-4.
FT  DISUFID 23 88 BY SIMILARITY.
FT  NON_TER 108
SQ  SEQUENCE 108 AA; 11850 MW; C5C145DC376F30CD CRC64;

Query Match 100.0%; Score 35; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 TDFTLTI 7
    |||||
DB  69 TDFTLTI 75

RESULT 13
KVIT_RAT
ID  KVIT_RAT      STANDARD:      PRT:  109 AA.
AC  P01681;
DT  21-JUL-1986 (Rel. 01, Created)
DT  21-JUL-1986 (Rel. 01, Last sequence update)
DT  15-JUL-1999 (Rel. 38, Last annotation update)
DE  Ig kappa chain V region S211.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE.
RA  STRAIN=LOU;
RA  MEDLINE=75212238; PubMed=807630;
RA  Stancov V., Querinjean P.;
RT  "The primary structure of a rat kappa Bence Jones protein:
RT  phylogenetic relationships of V- and C-region genes."
RL  J. Immunol. 115:59-62(1975).
CC  -I- MISCELLANEOUS: THIS IS A BENICE-JONES PROTEIN.
DR  PIR; A01944; KVT21.
DR  InterPro; IPR003006; Ig_MHC.
DR  InterPro; IPR003596; Ig_V.
DR  Pfam; PF00047; Ig; 1.
DR  SMART; SM00406; IGV; 1.
KM  Immunoglobulin V region; Bence-Jones protein.
FT  DOMAIN 1 23 FRAMEWORK-1.
FT  DOMAIN 24 35 COMPLEMENTARITY-DETERMINING-1.
FT  DOMAIN 36 50 FRAMEWORK-2.
FT  DOMAIN 51 57 COMPLEMENTARITY-DETERMINING-2.
FT  DOMAIN 58 89 FRAMEWORK-3.
FT  DOMAIN 90 98 FRAMEWORK-4.
FT  DOMAIN 99 108 FRAMEWORK-4.
FT  NON_TER 109
SQ  SEQUENCE 109 AA; 11947 MW; A25B9B9F9D5CBAC6 CRC64;

Query Match 100.0%; Score 35; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 TDFTLTI 7
    |||||
DB  70 TDFTLTI 76

RESULT 14
KVIT_HUMAN
ID  KVIT_HUMAN      STANDARD:      PRT:  109 AA.

```

AC P01612: 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain V-1 region Mew.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=83081018; PubMed=6816713;  
 RA Bult C., Linke R.P.;  
 RT Primary structure of the variable part of an amyloidogenic  
 RT Bence-Jones protein (Mew). An unusual insertion in the third  
 RT hypervariable region of a human kappa-immunoglobulin light chain.;  
 RL Hoppe-Seyler's Z. Physiol. Chem. 363:1347-1358(1982).  
 CC -1- MISCELLANEOUS: ANOTHER FORM THAT LACKED RESIDUES 1-3 WAS ALSO  
 CC FOUND.  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN.  
 DR PIR: A01879; KIDNEY.  
 DR HSSP; P01730; IMIO.  
 DR InterPro: IPR003506; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_V.  
 DR Pfam: PF00047; Ig\_1.  
 DR SMART: SM00406; IgV\_1.  
 KW Immunoglobulin V region.  
 FT DOMAIN 1 23 FRAMEWORK-1.  
 FT DOMAIN 2 34 COMPLEMENTARITY-DETERMINING-1.  
 FT DOMAIN 3 49 FRAMEWORK-2.  
 FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.  
 FT DOMAIN 57 88 FRAMEWORK-3.  
 FT DOMAIN 89 98 COMPLEMENTARITY-DETERMINING-3.  
 FT DOMAIN 99 108 FRAMEWORK-4.  
 FT DISULFID 23 88 BY SIMILARITY.  
 FT NON\_TER 109  
 FT SEQUENCE 109 AA: 11870 MW: 66ABF451D55F5A0 CRC64;

Query Match 100.0%; Score 35; DB 1; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 0.35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
 |||||  
 DB 69 TDFTLTI 75

RESULT 15  
 KV3B\_HUMAN STANDARD; PRT; 109 AA.  
 AC P01620:  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain V-III region SIE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=82046598; PubMed=6794615;  
 RA Andrews D.W., Capra J.D.;  
 RT Amino acid sequence of the variable regions of light chains from two  
 RT idiotypically cross-reactive human IgM anti-gamma-globulins of the Wa  
 RT group.;  
 RL Biochemistry 20:5816-5822(1981).  
 CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGM WITH ANTI-GAMMA  
 CC GLOBULIN ACTIVITY.  
 CC PIR: A01892; K3H0ST.  
 DR HSSP; P01789; IMCP.  
 DR InterPro: IPR003006; Ig\_MHC.

DR InterPro: IPR003596; Ig\_V.  
 DR Pfam: PF00047; Ig\_1.  
 DR SMART: SM00406; IgV\_1.  
 KW Immunoglobulin V region.  
 FT DISULFID 23 89 BY SIMILARITY.  
 FT NON\_TER 109  
 FT SEQUENCE 109 AA: 11775 MW: 7689C3ECD646FFB4 CRC64;

Query Match 100.0%; Score 35; DB 1; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 0.35;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
 |||||  
 DB 70 TDFTLTI 76

Search completed: July 15, 2002, 13:23:32  
 Job time: 1448 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:22:44 ; Search time 172.49 seconds

(without alignments)  
7.020 Million cell updates/sec

Title: US-09-712-819A-5

Perfect score: 35

Sequence: 1 TDFTLTI 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP virus:\*  
16: SP bacteriaph:\*  
17: SP archaeap:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	107	4 Q9UL81	Q9UL81 homo sapien
2	35	100.0	107	4 Q96SA9	Q96SA9 mus musculu
3	35	100.0	107	11 Q9ER29	Q9ER29 mus musculu
4	35	100.0	108	4 Q9UL79	Q9UL79 homo sapien
5	35	100.0	108	4 Q9UL77	Q9UL77 homo sapien
6	35	100.0	109	4 Q9UL70	Q9UL70 homo sapien
7	35	100.0	109	4 Q9UL86	Q9UL86 homo sapien
8	35	100.0	109	4 Q9UL78	Q9UL78 homo sapien
9	32	91.4	109	6 Q9N0W5	Q9N0W5 oryctolagus
10	32	91.4	16	Q99XR2	Q99XR2 streptococc
11	31	88.6	99	11 Q9UL74	Q9UL74 mus musculu
12	31	88.6	108	4 Q9UL83	Q9UL83 homo sapien
13	31	88.6	109	4 Q9UL85	Q9UL85 homo sapien
14	31	88.6	110	9 Q94MG5	Q94MG5 bacterioph
15	31	88.6	110	9 Q94MC3	Q94MC3 bacterioph
16	31	88.6	110	9 Q94MC2	Q94MC2 bacterioph

17	31	88.6	110	9 Q94MC1	Q94MC1 bacterioph
18	31	88.6	110	9 Q94MB9	Q94MB9 bacterioph
19	31	88.6	110	9 Q94LZ7	Q94LZ7 bacterioph
20	31	88.6	110	9 Q94LZ6	Q94LZ6 bacterioph
21	31	88.6	110	9 Q94LZ5	Q94LZ5 bacterioph
22	31	88.6	110	9 Q94LZ4	Q94LZ4 bacterioph
23	31	88.6	110	9 Q94LZ3	Q94LZ3 bacterioph
24	31	88.6	110	9 Q94LZ2	Q94LZ2 bacterioph
25	31	88.6	110	9 Q94LZ1	Q94LZ1 bacterioph
26	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
27	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
28	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
29	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
30	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
31	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
32	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
33	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
34	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
35	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
36	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
37	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
38	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
39	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
40	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
41	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
42	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
43	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
44	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph
45	31	88.6	110	9 Q94LZ0	Q94LZ0 bacterioph

## ALIGNMENTS

RESULT	ID	Q9UL81	PRELIMINARY:	PRT:	107 AA.
AC	Q9UL81	01-MAY-2000 (Trembl)	13, Created)		
DT	01-MAY-2000 (Trembl)	13, Last sequence update)			
DT	01-DEC-2001 (Trembl)	19, Last annotation update)			
DE	MYOSIN-REACTIVE IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION (FRAGMENT).				
OS	Homo sapiens (human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=98277139; PubMed=9614934;				
RA	Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,				
RA	Young D.C.;				
RT	*Myosin-reactive autoantibodies in rheumatic carditis and normal				
RT	fetus.*;				
RL	CLIN. Immunol. Immunopathol. 87:184-192(1998).				
DR	EMBL; AF035033; AAD56269.1; -.				
DR	HSSP; P01607; IREI.				
DR	InterPro; IPR003006; IG_MHC.				
DR	InterPro; IPR003596; IG_V.				
DR	Pfam; PF00047; IG_1.				
DR	SMART; SM00406; IG; 1.				
FT	NON_TER	1			
FT	NON_TER	107			
SO	SEQUENCE	107 AA;	11501 MW;	070549PDP0754748 CRC64;	

Query Match	100.0%;	Score 35;	DB 4;	Length 107;
Best Local Similarity	100.0%;	Pred. No. 3;		
Matches	7;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;
QY	1 TDFTLTI 7			
DB	69 TDFTLTI 75			

```

RESULT 2
ID Q96SA9 PRELIMINARY; PRT; 107 AA.
AC Q96SA9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ANTI-STREPTOCOCCAL/ANTI-MYOSIN IMMUNOGLOBULIN KAPPA LIGHT CHAIN
DE VARIABLE REGION (FRAGMENT).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98375893; PubMed=9712075;
RA Adderson E.E., Shikhan A.R., Ward K.E., Cunningham M.W.;
RT "Molecular analysis of polyclonal monoclonal antibodies from
RT rheumatic carditis: human anti-N-acetylglucosamine/anti-myosin
RT antibody v region genes";
RL T. Immunol. 161:2028-2031(1998).
DR EMBL; U96396; A868785.1; -.
FT NON_TER 1
FT NON_TER 107
SO SEQUENCE 107 AA; 11520 MW; 4BB4369C5B577F16 CRC64;

Query Match 100.0%; Score 35; DB 4; Length 107;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7
Db 69 TDFTLTI 75

RESULT 3
ID Q9ER29 PRELIMINARY; PRT; 107 AA.
AC Q9ER29;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ANTI HUMAN TNF-ALPHA LIGHT CHAIN VARIABLE REGION (FRAGMENT).
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen P., Deng J.B., Wang Z.L., Han H., Su C.Z.;
RT "Cloning and sequencing of the light chain fragment of variable region
RT genes of an anti-TNF- $\alpha$  monoclonal antibody.";
RL J. Cell. Mol. Immunol. 12:21-26(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Chen P., Deng J.B., Wang Z.L., Han H., Yao L.B., Su C.Z.;
RT "Construction and sequencing of the single-chain antibody gene of a
RT human TNF- $\alpha$  specific monoclonal antibody.";
RL Li 4 Chun 1 Ta Hsueh Hsueh Pao 19:373-376(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Chen P., Deng J.B., Wang Z.L., Han H., Yao L.B., Su C.Z.;
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF262753; AAG23804.1; -.
DR HSSP; P80362; 1WTL
DR InterPro; IPR003599; 1G.
DR InterPro; IPR003506; 1G_MHC.
DR InterPro; IPR003596; 1G_v.
DR Pfam; PF00047; 1G; 1.
DR SMART; SM00409; 1G; 1.

```

```

DR SMART; SM00406; 1Gv; 1.
FT NON_TER 107
SO SEQUENCE 107 AA; 11784 MW; 2B15FEA6604A26C3 CRC64;

Query Match 100.0%; Score 35; DB 11; Length 107;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7
Db 72 TDFTLTI 78

RESULT 4
ID Q9UL79 PRELIMINARY; PRT; 108 AA.
AC Q9UL79;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MYOSIN-REACTIVE IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION
DE (FRAGMENT).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.L., Kalls N.N., Berney S.M.,
RA Young D.C.;
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus.";
RL Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL; AF035035; AAD56271.1; -.
DR HSSP; P01607; 1KET.
DR InterPro; IPR003506; 1G_MHC.
DR InterPro; IPR003596; 1G_v.
DR Pfam; PF00047; 1G; 1.
DR SMART; SM00406; 1Gv; 1.
FT NON_TER 1
FT NON_TER 108
SO SEQUENCE 108 AA; 11787 MW; DB5845F19724FB4E CRC64;

Query Match 100.0%; Score 35; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7
Db 69 TDFTLTI 75

RESULT 5
ID Q9UL77 PRELIMINARY; PRT; 108 AA.
AC Q9UL77;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MYOSIN-REACTIVE IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION
DE (FRAGMENT).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98277139; PubMed=9614934;
RA Wu X., Liu B., Van der Merwe P.L., Kalls N.N., Berney S.M.,
RA Young D.C.;

```

RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 fetus."  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL: AF035037; AAD56273.1; -.  
 DR HSSP: P01607; 1REI.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_v.  
 DR Pfam: PF00047; Ig\_1.  
 DR SMART: SM00406; IgV; 1.  
 FT NON\_TER 1  
 FT NON\_TER 108  
 SQ SEQUENCE 108 AA; 11738 MW; C06681716CAD16F3 CRC64;

Query Match 100.0%; Score 35; DB 4; Length 108;  
 Best Local Similarity 100.0%; Pred. No. 3.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDEFTLT 7  
 Db 69 TDEFTLT 75

RESULT 6  
 Q9UL70 PRELIMINARY; PRT; 108 AA.  
 AC Q9UL70;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE MYOSIN-REACTIVE IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION  
 DE (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
 RA Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 fetus."  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL: AF035044; AAD56280.1; -.  
 DR HSSP: P01607; 1REI.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_v.  
 DR Pfam: PF00047; Ig\_1.  
 DR SMART: SM00406; IgV; 1.  
 FT NON\_TER 1  
 FT NON\_TER 108  
 SQ SEQUENCE 108 AA; 11633 MW; B7HEDC3E41FCCA37 CRC64;

Query Match 100.0%; Score 35; DB 4; Length 108;  
 Best Local Similarity 100.0%; Pred. No. 3.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDEFTLT 7  
 Db 69 TDEFTLT 75

RESULT 7  
 Q9UL86 PRELIMINARY; PRT; 109 AA.  
 AC Q9UL86;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE MYOSIN-REACTIVE IMMUNOGLOBULIN KAPPA CHAIN VARIABLE REGION  
 DE (FRAGMENT).

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
 RA Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 fetus."  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL: AF035028; AAD56264.1; -.  
 DR HSSP: P80362; 1WTL.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_v.  
 DR Pfam: PF00047; Ig\_1.  
 DR SMART: SM00406; IgV; 1.  
 FT NON\_TER 1  
 FT NON\_TER 109  
 SQ SEQUENCE 109 AA; 11928 MW; 243325F72C7DAC83 CRC64;

Query Match 100.0%; Score 35; DB 4; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 3.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDEFTLT 7  
 Db 70 TDEFTLT 76

RESULT 8  
 Q9UL78 PRELIMINARY; PRT; 109 AA.  
 AC Q9UL78;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE MYOSIN-REACTIVE IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION  
 DE (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,  
 RA Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 fetus."  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL: AF035036; AAD56272.1; -.  
 DR HSSP: P80362; 1WTL.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_v.  
 DR Pfam: PF00047; Ig\_1.  
 DR SMART: SM00406; IgV; 1.  
 FT NON\_TER 1  
 FT NON\_TER 109  
 SQ SEQUENCE 109 AA; 11646 MW; 5F679C52EC7EE197 CRC64;

Query Match 100.0%; Score 35; DB 4; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 3.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDEFTLT 7  
 Db 70 TDEFTLT 76

RESULT 9  
Q9NOW5 PRELIMINARY; PRT: 109 AA.  
ID Q9NOW5  
AC Q9NOW5  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ANTI-HUMAN A33 LIGHT CHAIN VARIABLE REGION (FRAGMENT).  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
[1]  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20250927; PubMed=10788485;  
RA Rader C., Rittler G., Nathan S., Elia M., Gout I., Jungbluth A.A.,  
RA Cohen L.S., Welt S., Old L.J., Barbas C.F. III.;  
RT "The rabbit antibody repertoire as a novel source for the generation  
RT of therapeutic human antibodies".  
RL J. Biol. Chem. 275:13668-13676(2000).  
DR EMBL: AF245502; AAF68449.1; -.  
DR HSSP: P80362; 1WTL.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003596; Ig\_V.  
DR Pfam: PF00047; 1g\_1.  
DR SMART: SM00406; IGV; 1.  
FT NON\_TER 1 109  
FT NON\_TER 1 109  
SQ SEQUENCE 109 AA; 11323 MW; BD8B36E75F947B CRC64;

Query Match 91.4%; Score 32; DB 6; Length 109;  
Best Local Similarity 85.7%; Pred. No. 14;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TDFTLTI 7  
Db 69 TDFTLTI 75  
RESULT 10  
Q9XR2 PRELIMINARY; PRT: 557 AA.  
ID Q9XR2  
AC Q9XR2  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PUTATIVE FORMATE-TETRAHYDROFOLATE LIGASE (EC 6.3.4.3).  
GN FHS.2 OR SPY2085.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
[1]  
RN  
RP SEQUENCE FROM N.A.  
RX STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;  
RX MEDLINE=21192684; PubMed=11296296;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lal H.S., Lin S.P.,  
RA Qian Y., Jia H.G., Nejar F.Z., Ren Q., Zhu R., Song L., Witte J.,  
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes".  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
DR EMBL: AE006628; AAK34738.1; -.  
DR InterPro: IPR000559; FTHFS.  
DR Pfam: PF01268; FTHFS; 1.  
DR PROSITE: PS00722; FTHFS\_2; 1.  
KW Ligase; Complete proteome.  
SQ SEQUENCE 557 AA; 59053 MW; CH07C9FCE90B3AE7 CRC64;

Query Match 91.4%; Score 32; DB 16; Length 557;  
Best Local Similarity 85.7%; Pred. No. 75;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TDFTLTI 7  
Db 505 TDFTLTI 511  
RESULT 11  
Q9JL74 PRELIMINARY; PRT: 99 AA.  
ID Q9JL74  
AC Q9JL74  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ANTI-MYOSIN IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION  
DE (FRAGMENT).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C;  
RX MEDLINE=20448942; PubMed=10992488;  
RA Maiktel S., Liao L., Cunningham M.W., Diamond B.;  
RT "T-Cell-dependent antibody response to the dominant epitope of  
RT streptococcal polysaccharide, N-acetyl-glucosamine, is cross-reactive  
RT with cardiac myosin".  
RL Infect. Immun. 68:5803-5808(2000).  
DR EMBL: AF206032; AAF69330.1; -.  
DR HSSP: P80362; 1WTL.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003596; Ig\_V.  
DR Pfam: PF00047; 1g\_1.  
DR SMART: SM00406; IGV; 1.  
FT NON\_TER 1 99  
FT NON\_TER 1 99  
SQ SEQUENCE 99 AA; 10939 MW; 3B25D0E78453324 CRC64;

Query Match 88.6%; Score 31; DB 11; Length 99;  
Best Local Similarity 85.7%; Pred. No. 21;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 TDFTLTI 7  
Db 61 TDFTLTI 67  
RESULT 12  
Q9UL83 PRELIMINARY; PRT: 108 AA.  
ID Q9UL83  
AC Q9UL83  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE MYOSIN-REACTIVE IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION  
DE (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98277139; PubMed=9614934;  
RA Wu X., Liu B., Van der Merwe P.L., Kalls N.N., Berney S.M.,  
RA Young D.C.;  
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
RT fetus".  
RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
DR EMBL: AF035031; AAD56267.1; -.  
DR HSSP: P80362; 1WTL.  
DR InterPro: IPR003006; Ig\_MHC.

DR InterPro: IPR003596; Ig\_v.  
 DR Pfam: PF00047; Ig; 1.  
 DR SMART: SM00406; IgV; 1.  
 FT NON\_TER 1  
 FT NON\_TER 108  
 SQ SEQUENCE 108 AA; 11834 MW; 9F9C5A92EBA96EEA CRC64;

Query Match 88.6%; Score 31; DB 4; Length 108;  
 Best Local Similarity 85.7%; Pred. No. 23;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
 |:|||||  
 Db 69 TDFTLTI 75

RESULT 13  
 ID Q9UL85 PRELIMINARY; PRT; 109 AA.  
 AC Q9UL85;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE MOSIN-REACTIVE IMMUNOGLOBULIN KAPPA CHAIN VARIABLE REGION  
 DE (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98277139; PubMed=9614934;  
 RA Wu X., Liu B., Van der Merwe P.L., Kallis N.N., Berney S.M.,  
 RA Young D.C.;  
 RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
 RT fetus.";  
 RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
 DR EMBL: AF035029; AAD56265.1; -  
 DR HSSP: P80362; 1WTL.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_v.  
 DR Pfam: PF00047; Ig; 1.  
 DR SMART: SM00406; IgV; 1.  
 FT NON\_TER 1  
 FT NON\_TER 109  
 SQ SEQUENCE 109 AA; 11761 MW; FB1E43E7C7AFACCC CRC64;

Query Match 88.6%; Score 31; DB 4; Length 109;  
 Best Local Similarity 85.7%; Pred. No. 24;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
 |:|||||  
 Db 69 TDFTLTI 75

RESULT 14  
 ID Q94MG5 PRELIMINARY; PRT; 110 AA.  
 AC Q94MG5;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PROTEIN GP30.8.  
 GN 30.8  
 OS Bacteriophage K3.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.  
 OX NCBI\_TaxID=10674;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Truncate L., Nivinskas R.;

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ315751; CAC42994.1; -  
 SQ SEQUENCE 110 AA; 12892 MW; 1685D7E297D83F7 CRC64;

Query Match 88.6%; Score 31; DB 9; Length 110;  
 Best Local Similarity 85.7%; Pred. No. 24;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
 |:|||||  
 Db 44 TDFTLTI 50

RESULT 15  
 ID Q94MG3 PRELIMINARY; PRT; 110 AA.  
 AC Q94MG3;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PROTEIN GP30.8.  
 GN 30.8  
 OS Bacteriophage Pol.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.  
 OX NCBI\_TaxID=36341;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Truncate L., Nivinskas R.; EMBL/GenBank/DBJ databases.  
 RL Submitted (JUN-2001) to the  
 DR EMBL: AJ315760; CAC43000.1; -  
 SQ SEQUENCE 110 AA; 12953 MW; B265DD59971389E6 CRC64;

Query Match 88.6%; Score 31; DB 9; Length 110;  
 Best Local Similarity 85.7%; Pred. No. 24;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
 |:|||||  
 Db 44 TDFTLTI 50

Search completed: July 15, 2002, 13:22:45  
 Job time: 1481 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 12:57:56 ; Search time 228.39 Seconds  
(without alignments)  
3.404 Million cell updates/sec

```
Title: US-09-712-819A-5
Perfect score: 35
Sequence: 1 TDFTLTI 7
```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

```
Searched:      747574 segs, 111073796 residues
Total number of hits satisfying chosen parameters:  747574
```

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database : A\_Geneseq.032802.\*

1:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1980.DAT *
2:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1981.DAT *
3:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1982.DAT *
4:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1983.DAT *
5:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1985.DAT *
6:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1986.DAT *
7:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1987.DAT *
8:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1988.DAT *
9:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1989.DAT *
10:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1990.DAT *
11:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1991.DAT *
12:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1992.DAT *
13:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1993.DAT *
14:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1994.DAT *
15:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1995.DAT *
16:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1996.DAT *
17:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1997.DAT *
18:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1998.DAT *
19:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA1999.DAT *
20:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA2000.DAT *
21:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA2001.DAT *
22:	/SIDS1/gcgdata/hold-genseq/genescqp-emb1/AA2002.DAT *

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	35	100.0	16	20	AAI41874	Rheumatoid arthritis
2	35	100.0	16	22	AAU25389	Schizophrenia-Asso
3	35	100.0	16	22	AAU15733	Schizophrenia-Asso
4	35	100.0	32	17	AAH87043	Human group 1 I-Igh
5	35	100.0	32	20	AA52745	Humanised Atr-5 L
6	35	100.0	32	22	AAH98286	Anti-A33 antigen i
7	35	100.0	32	22	AAH98288	Anti-A33 antigen i
8	35	100.0	32	22	AAH97666	A33 antigen bindin
9	35	100.0	32	22	AAH97668	A33 antigen bindin
10	35	100.0	74	19	AAH62805	Amino acid sequenc
11	35	100.0	76	20	AAH80981	Variable kappa I-Ig

12	35	100.0	82	19	AAW62807	Amino acid sequence
13	35	100.0	84	17	AAW14491	Monoclonal antibody
14	35	100.0	84	17	AAW98878	Monoclonal antibody
15	35	100.0	86	19	AAW62806	Amino acid sequence
16	35	100.0	86	19	AAW58619	Anti-RSV F protein
17	35	100.0	88	21	AAW56651	Partial peptide fragment
18	35	100.0	88	21	AAW56653	Partial peptide fragment
19	35	100.0	88	21	AAW56654	Partial peptide fragment
20	35	100.0	88	21	AAW56655	Partial peptide fragment
21	35	100.0	88	21	AAW56657	Partial peptide fragment
22	35	100.0	88	21	AAW56659	Partial peptide fragment
23	35	100.0	88	21	AAW56668	Partial peptide fragment
24	35	100.0	51	20	AAW95478	Mouse derived RT3
25	35	100.0	93	15	AAW54321	Anti-HIV gp120 immunogen
26	35	100.0	93	17	AAW01290	VL region of HIV gp120
27	35	100.0	93	17	AAW95142	Anti-gp120 antibody
28	35	100.0	93	21	AAW98251	Anti-gp120 antibody
29	35	100.0	94	19	AAW62808	Amino acid sequence
30	35	100.0	94	21	AAW56669	Partial peptide fragment
31	35	100.0	94	22	AAW67510	Light chain variable region
32	35	100.0	95	16	AAW72061	OF7K.11 VK-1 L chain
33	35	100.0	95	16	AAW72062	OF7K.9 VK-1 L chain
34	35	100.0	95	16	AAW72063	OF7K.17 VK-1 L chain
35	35	100.0	95	16	AAW72064	OF7K.17 VK-1 L chain
36	35	100.0	95	16	AAW72065	S43434 VK-1 region
37	35	100.0	95	16	AAW72058	KL012 VR region
38	35	100.0	95	16	AAW72059	OF7K.3 VR-1 L chain
39	35	100.0	95	16	AAW72060	OF7K.16 VK-1 L chain
40	35	100.0	100	13	AAW52524	LV region of human O
41	35	100.0	100	21	AAW64671	Human 5' EST related
42	35	100.0	101	13	AAW22577	Light chain VK10...
43	35	100.0	101	20	AAW34216	IgG antibody 2.1...
44	35	100.0	103	15	AAW56683	Humanized 359 light
45	35	100.0	103	15	AAW47933	Light chain region

## ALIGNMENTS

RESULT	1
AAV41874	AAV41874 standard; Peptide; 16 AA.
XX	
AC	AAV41874:
XX	
DT	09-DEC-1999 (first entry)
XX	
DE	Rheumatoid arthritis diagnostic protein isoform peptide #25.
XX	
KW	Human; rheumatoid arthritis; RA; diagnosis; RPI; RADP; detection;
KM	rheumatoid arthritis diagnostic feature; ERFI; synovial fluid;
KM	rheumatoid arthritis diagnostic protein isoform; screening;
XX	expression reference protein isoform; prognosis.
OS	Homo sapiens.
XX	
PN	WO947925-A2.
XX	
PD	23-SEP-1999.
XX	
PF	15-MAR-1999; 99WO-GB00763.
XX	
PR	13-MAR-1998; 98GB-0005477.
XX	
PA	(OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX	
PI	Parekh RB, Patel TP, Townsend RR;
XX	
DR	WPI; 1999-571871/48.
XX	
PT	Diagnosis of human rheumatoid arthritis by two-dimensional
XX	electrophoresis -
XX	

PS Disclosure: Page 18; 157pp: English.

XX A method has been developed for the diagnosis of human rheumatoid  
CC arthritis (RA) using two-dimensional electrophoresis to generate a  
CC two-dimensional array of features. The method can be used for screening,  
CC diagnosis and prognosis of RA in a subject or for monitoring the effect  
CC of an anti-RA drug or therapy administered to a subject. The method  
CC comprises: (a) analysing a sample of serum or plasma and optionally  
CC synovial fluid by two-dimensional electrophoresis, to generate a two-  
CC dimensional array of features; (b) identifying at least one chosen  
CC feature whose relative abundance correlates with the presence or absence  
CC of RA; and (c) comparing the abundance of each chosen feature in the  
CC sample with the abundance of that chosen feature in serum or plasma from  
CC one or more persons without RA, where the relative abundance of the  
CC chosen feature or features in the sample indicates the presence or  
CC absence of RA in the subject. The method can also be used in clinical  
CC studies for testing drugs for therapy of RA, for purification of RA-  
CC diagnostic protein isoforms (RPIs), and for production of antibodies to  
CC RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify  
CC compounds that promote or inhibit their activity, which are then used as  
CC RA drugs. Nucleic acid encoding RADFs can be used in gene therapy  
CC protocols. AAY41844 to AAY42100 represent RPI peptides. AAY42101 to  
CC AAY42506 to AAZ25068 represent reference protein isoform peptides and  
CC used in the exemplification of the present invention.

XX Sequence 16 AA:

Query Match 100.0%; Score 35; DB 20; Length 16;

Best Local Similarity 100.0%; Pred. No. 1;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
|||||||  
Db 8 tdfeltl 14

RESULT 2

AAU25389 standard; Peptide: 16 AA.

XX AAU25389;

DT 18-DEC-2001 (first entry)

DE Schizophrenia-Associated protein isoform (SPI) peptide #618.

XX Schizophrenia-associated protein isoform; SPI; SPI-206; SPI-238; SPI-240;

KM neuroleptic; gene therapy; cerebrospinal fluid; serum; plasma.

XX Homo sapiens.

PN WO200162785-A2.

PD 30-AUG-2001.

PF 23-FEB-2001; 2001WO-GB00792.

PR 24-FEB-2000; 2000GB-0004415.

PR 28-NOV-2000; 2000US-0750395.

XX (OXFO-) OXFORD GLYSCSCIENCES UK LTD.

PI Herath HMAc, Parekh RB, Rohlf C, Terrett JA, Tyson RL;

DR WPI: 2001-570624/64.

PT New schizophrenia associated protein isoforms and encoding nucleic acid  
PT molecules, useful for treatment, diagnosis and prognosis of  
PT schizophrenia and screening for potential drugs for treatment and new  
PT drug targets -

PS Disclosure: Page 41; 148pp: English.

XX The sequence represents a schizophrenia-associated protein isoform (SPI).  
CC These protein isoforms, e.g. SPI-206, SPI-228 and SPI-240 are detectable  
CC in cerebrospinal fluid, serum or plasma and are useful markers of  
CC schizophrenia. The sequences can be used for treatment and diagnosis of  
CC schizophrenia, screening, prognosis, monitoring the results of therapy,  
CC identifying patients most likely to respond to a particular therapy and  
CC identification of new targets for drug treatment. SPI DNA is useful as a  
CC nucleic acid probe to detect the presence of nucleic acids or SPIs.

XX Sequence 16 AA:

Query Match 100.0%; Score 35; DB 22; Length 16;

Best Local Similarity 100.0%; Pred. No. 1;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
|||||||  
Db 8 tdfeltl 14

RESULT 3

AAU15733 standard; Peptide: 16 AA.

XX AAU15733;

DT 24-OCR-2001 (first entry)

DE Schizophrenia-associated isoform peptide #618.

XX Schizophrenia: neuroleptic; diagnostic; neuropsychiatric disorder;

KW neurological disorder; neuropathy.

XX Homo sapiens.

PN WO200163293-A2.

PD 30-AUG-2001.

PF 23-FEB-2001; 2001WO-GB00783.

PR 24-FEB-2000; 2000GB-0004415.

PR 28-NOV-2000; 2000US-0750395.

XX (OXFO-) OXFORD GLYSCSCIENCES UK LTD.

PI Herath HMAc, Parekh RB, Rohlf C;

DR WPI: 2001-502868/55.

PT Diagnosing and monitoring schizophrenia by detecting the presence of  
PT schizophrenia associated features and schizophrenia associated protein  
PT isoforms in samples of cerebrospinal fluid -

XX Claim 6; Page 41; 160pp: English.

XX The invention relates to methods and compositions for screening,  
CC diagnosis and prognosis of Schizophrenia. The method involves detecting  
CC the presence of Schizophrenia (SCH) Associated Features (SFS) and SCH  
CC Associated Protein Isoforms (SPIs) in samples, e.g. by electrophoresis,  
CC immunoassay or hybridisation assay, for diagnosing and monitoring SCH,  
CC studying the effectiveness of treatments and for identifying potential  
CC therapeutic agents. The method is used for (1) screening or diagnosis of  
CC SCH and the relative abundance of at least 1 chosen feature correlates  
CC with the presence or absence of SCH; and (2) monitoring the effect of  
CC therapy administered to a subject with SCH and the relative abundance of  
CC at least 1 chosen feature which correlates with the severity of SCH.  
CC The expression and activity of the SFS, SPIs and related molecules  
CC (e.g. secondary messengers) are studied to diagnose SCH, monitor the  
CC progress of the disorder and the effectiveness of treatment and as



CC targets to identify and produce potential therapeutic agents for the  
 CC treatment of SCH. The paucity of detectable neuralgic defects  
 CC distinguishes neuropsychiatric disorders such as SCH from neurological  
 CC disorders, where manifestations of anatomical and biochemical changes  
 CC have been identified in many cases. Consequently the identification and  
 CC characterisation of cellular and/or molecular causative defects and  
 CC neuropathies are necessary for improved treatment of neuropsychiatric  
 CC disorders. AAU5114-AAU5162 represent the amino acid sequences of  
 CC schizophrenia-associated isoforms used in the method of the invention.

XX Sequence 16 AA;

Query Match 100.0%; Score 35; DB 22; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFTLTI 7  
 |||||  
 Db 8 tdfctli 14

#### RESULT 4

ID AAR87043 standard; Peptide: 32 AA.

AC AAR87043;

DT 25-JUN-1996 (first entry)

DE Human group I light chain framework 3.

XX Humanised antibody; interleukin-5; IL-5; recombinant antibody;

KW antibody engineering; monoclonal antibody; Mab; 39D10; CDR;

KM complementarity determining region; light chain; framework;

KW eosinophilia; allergy; asthma.

OS Homo sapiens.

PN WO9535375-A1.

PD 28-DEC-1995.

PF 16-JUN-1995; 95WO-GB01411.

PR 17-JUN-1994; 94GB-0012230.

PA (CLLT ) CELUTECH THERAPEUTICS LTD.

PI Athwal DS, Bodmer MM, Entage JS;

DR WPI; 1996-058412/06.

PT Anti-human IL-5 recombinant antibody - useful for preventing or

PT reducing eosinophilia and for treating certain allergic diseases,

PT esp. asthma

PS Example 3; Fig 3; 69pp; English.

XX Framework regions (AAR87041-44) of human group I (gpl) germ line

CC antibody light chain showed homology to corresponding regions-5

CC (AAR87045-48, respectively) of the rat anti-human interleukin-5

CC monoclonal antibody 39D10 light chain (see AAR87040). This homology

CC was utilised in the prodn. of a humanised 39D10 VL (AAR87057) in

CC which rat 39D10 VL complementarity determining regions were grafted

CC into the human gpl framework.

XX Sequence 32 AA;

Query Match 100.0%; Score 35; DB 17; Length 32;

Best Local Similarity 100.0%; Pred. No. 2.1; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0;

QY 1 TDFTLTI 7  
 |||||  
 Db 13 tdfctli 19

#### RESULT 5

ID AAY52745 standard; Peptide: 32 AA.

AC AAY52745;

DT 26-JAN-2000 (first entry)

DE Humanised ATR-5 L chain V region FR3 for "a".

XX Human tissue factor; TF; humanised; antibody; mouse monoclonal antibody;

KW ATR-2; ATR-3; ATR-4; ATR-5; ATR-7; ATR-8; thrombotic disease; DIC;

KM disseminated intravascular coagulation; immunogenicity; chimeric.

OS Synthetic.

PN WO951743-A1.

PD 14-OCT-1999.

PF 02-APR-1999; 99WO-JP01768.

PR 03-APR-1998; 98JP-0091850.

PA (CHUS ) CHUGAI SEIYAKU KK.

PI Sato K, Adachi H, Yabuta N;

DR WPI; 1999-620204/53.

PT Humanised antibody recognising human tissue factor, used for treatment

PT of disseminated intravascular coagulation

PS Claim 17; Page 270; 291pp; Japanese.

XX The present invention describes chimeric antibody (Ab) heavy (H) chains

CC containing the variable region of the H chain of a mouse monoclonal Ab

CC recognising human tissue factor (hTF) and the constant region of the H

CC chain of a human Ab. The variable region is one of six specified

CC sequences (which are the H chain variable regions from mouse monoclonal

CC Ab's ATR-2,3,4,5,7 or 8). Also described are chimeric Ab light (L)

CC chains containing the variable region of the L chain of a mouse

CC monoclonal Ab recognising human tissue factor (hTF) and the constant

CC region of the L chain of a human Ab, the variable region being one of six

CC specified sequences (which are the L chain variable regions from mouse

CC monoclonal Ab's ATR-2,3,4,5,7 or 8). The chimeric Ab's can be used for

CC the treatment and prevention of thrombotic disease, especially of

CC disseminated intravascular coagulation (DIC). The humanised antibody has

CC the high hTF binding activity of the mouse monoclonal antibody but

CC greatly reduced immunogenicity. AAZ33001 to AAZ33091 and Y527007 to

CC AAY52767 represent sequences used in the exemplification of the present

CC invention.

XX Sequence 32 AA;

Query Match 100.0%; Score 35; DB 20; Length 32;

Best Local Similarity 100.0%; Pred. No. 2.1; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0;

QY 1 TDFTLTI 7  
 |||||  
 Db 13 tdfctli 19

#### RESULT 6

ID	AA98286
AC	AA98286; standard; Peptide: 32 AA.
DT	20-AUG-2001 (first entry)
DE	Anti-A33 antigen immunoglobulin VL FR3 SEQ ID NO:92.
KW	Human; rabbit; humanised; A33 antigen; anti-A33 antigen antibody;
KM	immunoglobulin; complementarily determining region; CDR; cancer;
KX	cytostatic; anticancer; colon cancer; stomach cancer.
OS	Homo sapiens.
PN	MO200130393-AZ.
PD	03-MAY-2001.
PE	20-OCT-2000; 2000WO-US29289.
PR	22-OCT-1999; 99US-0425638.
PR	04-APR-2000; 2000US-0543004.
PA	(LUDW-) LUDWIG INST CANCER RES.
PI	(SLOK) SLOAN KETTERING INST CANCER RES.
PJ	(SCRI) SCRIPPS RES INST.
XX	Barbas CF, Rader C, Ritter G, Welt S, Old LJ;
XX	WPI: 2001-328613/34.
PT	Treating cancers, particularly of stomach and colon, that express A33
PT	antigen by administering conjugates of anticancer agent with specific
PT	immunoglobulin product -
XX	
PS	Claim 16; Page 40; 85pp; English.
CC	The present invention describes a method for treating cancers that
CC	express the A33 antigen. The method comprises administering an
CC	anticancer agent (I) conjugated to an immunoglobulin product (II) that
CC	binds specifically to A33 and contains one or more of 13 specified
CC	complementarily determining regions (CDRs), given in AA98286 to
CC	AA98274. (I) has cytostatic activity. The method can be used for
CC	treating colon and stomach cancers. (II), or the nucleic acid encoding
CC	it, can be used directly, in unconjugated form, for immunotherapy of
CC	cancer, and, when labeled, for detection or diagnosis of diseases
CC	associated with A33 expression. AAH22218 to AAH22254 and AA98230 to
CC	AA98321 represent sequences used in the exemplification of the
CC	present invention.
XX	
SQ	Sequence 32 AA:
Query Match	100.0%; Score 35; DB 22; Length 32;
Best Local Similarity	100.0%; Pred. No. 2.1;
Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 TDFLTIT 7
DB	13 tdfltitl 19
RESULT	7
ID	AA98288
AC	AA98288 standard; Peptide: 32 AA.
XX	
XX	AA98288;
DT	20-AUG-2001 (first entry)
DE	Anti-A33 antigen immunoglobulin VL FR3 SEQ ID NO:94.
XX	

KW	Human; antibody; humanised; A33 antigen; anti-A33 antigen antibody;
KV	Immunoglobulin; complementarily determining region; CDR; cancer;
KW	Cytostatic; anticancer; colon cancer; stomach cancer.
XX	
OS	Homo sapiens.
PM	WO200130393-A2.
XX	
PD	03-MAY-2001.
XX	
PF	20-OCT-2000; 2000MO-US29289.
XX	
PR	22-OCT-1999; 99US-0425638.
PR	04-APR-2000; 2000US-0543004.
XX	
PA	(LUDM-) LUDMG INST CANCER RES.
PA	(SLOC) SLOAN KETTERING INST CANCER RES.
PA	(SCRI ) SCRIPPS RES INST.
P1	Barbas CF, Rader C, Rittter G, Well S, Old LJ;
XX	
DR	WPI: 2001-328613/34.
PT	Treating cancers, particularly of stomach and colon, that express A33
PT	antigen by administering conjugate of anticancer agent with specific
PT	immunoglobulin product -
XX	
ES	Claim 16; Page 40; 85pp; English.
XX	
CC	The present invention describes a method for treating cancers that
CC	express the A33 antigen. The method comprises administering an
CC	anticancer agent (i) conjugated to an immunoglobulin product (ii) that
CC	binds specifically to A33 and contains one or more of 13 specified
CC	complementarily determining regions (CDRs), given in AAB98262 to
CC	AAB98277. (I) has cytostatic activity. The method can be used for
CC	treating colon and stomach cancers. (II), or the nucleic acid encoding
CC	it, can be used directly, in unconjugated form, for immunotherapy of
CC	cancer, and, when labeled, for detection or diagnosis of diseases
CC	associated with A33 expression. AAH22218 to AAH22254 and AAB98230 to
CC	AAB98321 represent sequences used in the exemplification of the
CC	present invention.
XX	
SQ	Sequence 32 AA:
Query Match	100.0%; Score 35; DB 22; Length 32;
Best Local Similarity	100.0%; Pred. No. 2,1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 TDEFTLT I 7
	I I I I I I I I
Db	13 tdfItLI 19
RESULT 8	
ID	AAB97666 standard; Peptide; 32 AA.
XX	
AC	AAB97666;
XX	
DT	08-AUG-2001 (first entry)
XX	
DE	A33 antigen binding immunoglobulin product VLEF3 peptide SEQ ID NO:92.
XX	
KW	Chimeric antibody; humanised; humanisation; mammalian antibody; antigen;
KW	immunoreact; anti A33 antigen antibody; immunoglobulin.
XX	
OS	Homo sapiens.
XX	
PN	WO200131065-A1.
XX	
PD	03-MAY-2001.
XX	

PF 20-OCT-2000; 2000WO-US29026.  
 XX  
 PR 22-OCT-1999; 99US-0425638.  
 PR 04-APR-2000; 2000US-0543004.  
 XX  
 PA (SCRI ) SCRIPPS RES INSTR.  
 XX  
 PI Barbass CF, Rader C;  
 XX  
 DR WPI; 2001-328657/34.  
 XX  
 PR Preparing humanized rabbit antibodies that specifically immunoreact  
 PR with a particular antigen using display technology for expressing  
 PR libraries of antibody domains and fine tuning variable domain regions -  
 XX  
 XX  
 PS Example 9; Page 39; 62pp; English.  
 XX  
 CC The present invention describes a method for preparing a humanised rabbit  
 CC antibody that specifically immunoreacts with a particular antigen. The  
 CC method comprises expressing a library of antibodies comprising one or  
 CC more complementarily determining region (CDR) from the variable domain  
 CC sequences that specifically immunoreact with the antigen grafted into  
 CC framework regions from humans, and selecting the antibodies that react  
 CC with the antigen. The method is useful for humanising non-human  
 CC mammalian antibodies, which can be used for the treatment of a variety  
 CC of diseases. The present sequence represents an A33 antigen binding  
 CC immunoglobulin product VLFR3 peptide which is given in an example from  
 CC the present invention.  
 XX  
 SQ Sequence 32 AA:

Query Match 100.0%; Score 35; DB 22; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
 |||||  
 Db 13 tdfitli 19

RESULT 9  
 ID AAB97668 standard; Peptide; 32 AA.  
 XX  
 AC AAB97668;  
 XX  
 DT 08-AUG-2001 (first entry)  
 XX  
 DE A33 antigen binding immunoglobulin product VLFR3 peptide SEQ ID NO:94.  
 XX  
 KW Chimeric antibody; humanised; humanisation; mammalian antibody; antigen;  
 KW immunoreact; anti A33 antigen antibody; immunoglobulin.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200131065-A1.  
 XX  
 PD 03-MAY-2001.  
 XX  
 PR 20-OCT-2000; 2000WO-US29026.  
 XX  
 PR 22-OCT-1999; 99US-0425638.  
 PR 04-APR-2000; 2000US-0543004.  
 XX  
 PA (SCRI ) SCRIPPS RES INSTR.  
 XX  
 PI Barbass CF, Rader C;  
 XX  
 DR WPI; 2001-328657/34.  
 XX  
 PR Preparing humanized rabbit antibodies that specifically immunoreact  
 PR with a particular antigen using display technology for expressing

PF libraries of antibody domains and fine tuning variable domain regions -  
 XX  
 PS Example 9; Page 39; 62pp; English.  
 XX  
 CC The present invention describes a method for preparing a humanised rabbit  
 CC antibody that specifically immunoreacts with a particular antigen. The  
 CC method comprises expressing a library of antibodies comprising one or  
 CC more complementarily determining region (CDR) from the variable domain  
 CC sequences that specifically immunoreact with the antigen grafted into  
 CC framework regions from humans, and selecting the antibodies that react  
 CC with the antigen. The method is useful for humanising non-human  
 CC mammalian antibodies, which can be used for the treatment of a variety  
 CC of diseases. The present sequence represents an A33 antigen binding  
 CC immunoglobulin product VLFR3 peptide which is given in an example from  
 CC the present invention.  
 XX  
 SQ Sequence 32 AA:

Query Match 100.0%; Score 35; DB 22; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
 |||||  
 Db 13 tdfitli 19

RESULT 10  
 ID AAW62805 standard; Peptide; 74 AA.  
 XX  
 AC AAW62805;  
 XX  
 DT 23-SEP-1998 (first entry)  
 XX  
 DE Amino acid sequence of a human antibody fragment.  
 XX  
 KW Human; immunoglobulin; Ig; transgenic; non-human mammal;  
 KW inactivated endogenous Ig locus; B-cell development;  
 KW human heavy chain Ig locus; micro constant region; J-H; D-H; V-H gene;  
 KW kappa light chain Ig locus; kappa constant region; J-kappa gene; V-kappa;  
 KW production; antibody.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9824893-A2.  
 XX  
 PD 11-JUN-1998.  
 XX  
 PR 03-DEC-1997; 97WO-US23091.  
 XX  
 PR 03-DEC-1996; 96US-0759620.  
 XX  
 PA (ABGE-) ABGENIX INC.  
 XX  
 PI Green L, Jakobovits A, Klapholz S, Kucherlapati R;  
 PI Mendez M;  
 XX  
 DR WPI; 1998-333314/29.  
 XX  
 PR New transgenic non-human mammals - having an inactivated  
 PR immunoglobulin locus and a near complete human immunoglobulin locus,  
 PR used for production of human antibodies  
 XX  
 PS Disclosure; Page 77; 128pp; English.  
 XX  
 CC AAW62793-822 represent fragments of human antibodies produced by  
 CC transgenic xenomice, created using the method of the invention. The  
 CC specification describes a transgenic non-human mammal which has genome  
 CC modifications that comprise an inactivated endogenous immunoglobulin (Ig)  
 CC locus, so that the mammal does not display normal B-cell development. The  
 CC modified genome also has an inserted human heavy chain Ig locus in

CC germ-line configuration, the human heavy chain Ig locus comprising a human  
CC micro constant region and regulatory and switch sequences, human J-H  
CC genes, human D-H genes, and human V-H genes and an inserted human kappa  
CC light chain Ig locus in germ-line configuration, the human kappa light  
CC chain Ig locus comprising a human kappa constant region, J-kappa genes,  
CC and V-kappa genes, where the number of V-H and V-kappa genes inserted  
CC are selected to restore normal B-cell development in the mammal. The  
CC transgenic animals have a near complete human Ig locus, including both a  
CC human heavy chain locus and a human kappa light chain locus. They can  
CC be used for the production of human antibodies when exposed to  
CC particular antigens e.g. When exposed to human IL-8, EGFR or TNF- alpha  
CC the mice will produce antibodies to IL-8, EGFR or TNF- alpha  
CC respectively.

Query Match	100.0%	Score 35;	DB 19;	Length 74;
Best Local Similarity	100.0%	Pred. No. 5.1;		
Matches	7;	Conservative	0;	Mismatches
			0;	Indels
				0;
QY	1 TDFTLTI	7		
Db	48 tdfctli	54		

RESULT	11
AAW80981	
ID	AAW80981 standard; Protein; 76 AA.

AC AAW80981;

DT 30-MAR-1999 (first entry)

Variable kappa light region 012 encoded amino acid.

KW Human; epidermal growth factor receptor; tumour; EGF;  
KW transforming growth factor alpha; TGF-alpha.

OS Homo sapiens.

PN W09850433-Å2.

PD 12-NOV-1998.

PF 05-MAY-1998; 98WO-US09160.

PR 05-MAY-1997; 97US-0851362.

PA (ABGE-) ABGENIX INC.

PI Gallo M, Jakobovits A, Ula X, Yang X;

WPI; 1999-034712/03.

Example 3, Page 105; 143pp; English.

The variable kappa light region 012 encoded amino acid was used in the production of anti-epidermal growth factor receptor (EGF-r)-antibodies. The antibodies can be administered therapeutically to patients (human or veterinary) to treat solid tumours. EGF-r is overexpressed on many human solid tumour types, and the fully human antibodies (i.e. comprising and inhibiting both epidermal growth factor (EGF) and transforming growth factor alpha (TGF- $\alpha$ ) binding to EGF-r (known to lead to cellular proliferation and tumour growth). They can prevent tumour cell growth and, in combination with an antineoplastic agent, may eradicate established tumours. The fully human antibodies can minimise the immunogenic and allergic responses intrinsic to previous mouse/rat or mouse/rat-derived antibodies.

XX	Sequence	76 AA;
SQ		

Query Match 100.0%; Score 35; DB 20; Length 76;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TDFTLTI	7
Db	50	tdftlti	56

```

RESULT 12
AAW62807
ID AAW62807 standard; Peptide; 82 AA

```

AC AAW62807;

DT 23-SEP-1998 (first entry)

Amino acid sequence of a human antibody fragment.

KM Human; immunoglobulin; Ig; transgenic; non-human mammal;  
 KM inactivated endogenous Ig locus; B-cell development;  
 KM human heavy chain Ig locus; micro constant region; J-H; D-H; V-H gene;  
 KM kappa light chain Ig locus; kappa constant region; J-kappa gene; V-kappa  
 KM production; antibody.

OS Homo sapiens.

PN W09824893-A2.

PD 11-JUN-1998

PF 03-DEC-1997; 97WO-US23091.

PR 03-DEC-1996; 96US-0759620.

PA (ABGE-) ABGENIX INC.

PI Green L, Jakobovits A, Klapholz S, Kucherlapati R;

PI Méndez M;

DR WPI; 1998-333314/29.

PT New transgenic non-human mammals - having an inactivated immunoglobulin locus and a near complete human immunoglobulin locus, used for production of human antibodies

PS Disclosure; Page 78; 128pp; English

AA62793322 represent fragments of human antibodies produced by transgenic Xenopus, created using the method of the invention. The specification describes a transgenic non-human mammal which has genome modifications that comprise an inactivated endogenous immunoglobulin (Ig) locus, so that the mammal does not display normal B-cell development. The modified genome also has an inserted human heavy chain Ig locus in germline configuration, the human heavy chain Ig locus comprising a micro constant region, the human heavy chain Ig locus comprising a human D-H genes, and human V-H genes and switch sequences, human J-H genes, human D-H genes, and human V-H genes and an inserted human kappa light chain Ig locus comprising a human kappa constant region, J-kappa genes, and V-kappa genes, where the number of V-H and V-kappa genes inserted are selected to restore normal B-cell development in the mammal. The transgenic animals have a near complete human Ig locus, including both a human heavy chain locus and a human kappa light chain locus. They can be used for the production of human antibodies when exposed to particular antigens e.g. when exposed to human IL-8, EGFR or TNF- alpha the mice will produce antibodies to IL-8, EGFR or TNF- alpha respectively.

SQ Sequence 82 AA;

Query Match 100.0%; Score 35; DB 19; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 5.7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFLTIT 7  
 |||||  
 Db 56 tdfilti 62

## RESULT 13

AAW14491  
 ID AAW14491 standard; Protein; 84 AA.

AC AAW14491;

DE 28-JAN-1997 (first entry)

XX Monoclonal antibody D VK.

XX heavy; light chain; monoclonal antibody; antigen 3; marker; melanoma;

KW permanent human tumour cell line; tumour-associated antigen; epitope;

KW gastrointestinal tumour; pancreatic carcinoma; diagnostic; therapeutic;

XX antigen 11; Vibrio cholera; neuraminidase-resistant; ganglioside GD2.

XX Synthetic.

XX EP727436-A1.

XX 21-AUG-1996.

XX 21-MAR-1990; 90EP-0105322.

XX 24-MAR-1989; 89DE-3909799.

PA (BEHW ) BEHRINGWERKE AG.

PI Auerbach B, Bosslet K, Sedlacek H, Seemann G;

XX WPI: 1996-372835/38.

DR N-PSDB; AAT63508.

PT Monoclonal antibody to tumour-associated antigen - useful as

XX gastrointestinal tumour marker

PS Disclosure; Page 14; 19pp; German.

XX AAW14490-91 are the heavy and light chains (respectively) of monoclonal

CC antibody (Mab) D. Mab D recognises Vibrio cholera

CC neuraminidase-resistant epitope of ganglioside GD2, from a human melanoma

CC cell line. Mabs A, B and C (see AAW1484-89) are mentioned in the

CC specification, but are not part of the claims. Mabs A and B recognise

CC antigens 3 and 11 resp. of a permanent human tumour cell line. Mab C

CC also recognises an epitope of a tumour-associated antigen. These antigens

CC occur at high concns. In the serum of patients with gastrointestinal

CC tumours, e.g. pancreatic carcinoma, and are thus useful as tumour markers

CC for diagnostic or therapeutic purposes.

XX Sequence 84 AA;

Query Match 100.0%; Score 35; DB 17; Length 84;

Best Local Similarity 100.0%; Pred. No. 5.8;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFLTIT 7  
 |||||

Db 47 tdfilti 53

## RESULT 14

AA99878

ID AAR99878 standard; Protein; 84 AA.

AC AAR99878;

DE 28-JAN-1997 (first entry)

XX Monoclonal antibody D VK.

KW Monoclonal antibody; Mab; epitope; tumour-associated antigen;

XX marker; antigen.

XX Synthetic.

XX EP727435-A1.

XX 21-AUG-1996.

XX 21-MAR-1990; 90EP-0105322.

XX 24-MAR-1989; 89DE-3909799.

PA (BEHW ) BEHRINGWERKE AG.

PI Auerbach B, Bosslet K, Sedlacek H, Seemann G;

XX WPI: 1996-372835/38.

DR N-PSDB; AAT36666.

PT Monoclonal antibody to tumour-associated antigen - useful as

XX gastrointestinal tumour marker

PS Disclosure; Page 14; 19pp; German.

XX Mab C (AAT36659-736660) is a monoclonal antibody that recognises an

CC epitope of a tumour-associated antigen occurring at high concn. In

CC the serum of patients with gastrointestinal tumours, e.g. pancreatic

CC carcinoma, and is thus useful as a tumour marker for diagnostic or

CC therapeutic purposes.

CC Mabs A, B and D are mentioned in the specification, but are not

CC part of the claims.

CC Mab A (AAT36661-736662) recognises antigen 3 of permanent human

CC tumour cell line.

CC Mab B (AAT36663-736664) recognises antigen 11 of permanent human

CC tumour cell line.

CC Mab D (AAT36665-736666) recognises a Vibrio cholera neuraminidase-

CC resistant epitope of ganglioside GD2, from a human melanoma cell

CC line.

XX Sequence 84 AA;

Query Match 100.0%; Score 35; DB 17; Length 84;

Best Local Similarity 100.0%; Pred. No. 5.8;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 15

AAW62806  
 ID AAW62806 standard; Peptide; 86 AA.

AC AAW62806;

DE 23-SEP-1998 (first entry)

XX Amino acid sequence of a human antibody fragment.

XX Human: immunoglobulin; Ig; transgenic; non-human mammal;

KW inactivated endogenous Ig locus; B-cell development;

KW human heavy chain Ig locus; micro constant region; J-H; D-H; V-H gene;

KM kappa light chain Ig locus; kappa constant region; J-kappa gene; V-kappa;  
 KW production; antibody.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09824893-A2.  
 XX  
 PD 11-JUN-1998.  
 XX  
 PF 03-DEC-1997; 97NO-US23091.  
 XX  
 PR 03-DEC-1996; 96US-0759620.  
 XX  
 PA (ABGE-) ABGENIX INC.  
 XX  
 PI Green L, Jakobovits A, Klapholz S, Kucherlapati R;  
 PI Mendez M;  
 DR WPI: 1998-333314/29.  
 XX  
 PT New transgenic non-human mammals - having an inactivated  
 PT immunoglobulin locus and a near complete human immunoglobulin locus,  
 PT used for production of human antibodies  
 XX  
 PS Disclosure; Page 78; 128pp; English.  
 XX  
 CC AAW62793-822 represent fragments of human antibodies produced by  
 CC transgenic xenotice, created using the method of the invention. The  
 CC specification describes a transgenic non-human mammal which has genome  
 CC modifications that comprise an inactivated endogenous immunoglobulin (Ig)  
 CC locus, so that the mammal does not display normal B-cell development. The  
 CC modified genome also has an inserted human heavy chain Ig locus in  
 CC germline configuration, the human heavy chain Ig locus comprising a human  
 CC micro constant region and regulatory and switch sequences, human J-H  
 CC genes, human D-H genes, and human V-H genes and an inserted human kappa  
 CC light chain Ig locus in germline configuration, the human kappa light  
 CC chain Ig locus comprising a human kappa constant region, J-kappa genes,  
 CC and V-kappa genes, where the number of V-H and V-kappa genes inserted  
 CC are selected to restore normal B-cell development in the mammal. The  
 CC transgenic animals have a near complete human Ig locus, including both a  
 CC human heavy chain locus and a human kappa light chain locus. They can  
 CC be used for the production of human antibodies when exposed to  
 CC particular antigens e.g. when exposed to human IL-8, EGFR or TNF- alpha  
 CC the mice will produce antibodies to IL-8, EGFR or TNF- alpha  
 CC respectively.  
 CC  
 XX  
 SQ Sequence 86 AA:  
 XX

Query Match 100.0%; Score 35; DB 19; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 6;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TDFRTITI 7  
 |||||||  
 Db 48 tdfrtlti 54

Search completed: July 15, 2002, 12:57:57  
 Job time: 413 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 12:59:18 ; Search time 75.67 Seconds  
(without alignments)  
2.260 Million cell updates/sec

Title: US-09-712-819A-5  
Perfect score: 35  
Sequence: 1 TDFTLTI 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCNUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	100.0	25	5	PCT-US91-02942-42
2	35	100.0	32	2	US-08-470-139-13
3	35	100.0	32	4	US-08-569-147-40
4	35	100.0	32	4	US-09-347-061-13
5	35	100.0	32	4	US-09-425-638A-92
6	35	100.0	32	4	US-09-425-638A-94
7	35	100.0	32	4	US-09-543-004-92
8	35	100.0	32	4	US-09-543-004-94
9	35	100.0	50	5	PCT-US91-02942-9
10	35	100.0	52	1	US-08-162-102C-43
11	35	100.0	53	1	US-08-162-102C-44
12	35	100.0	64	2	US-08-765-179B-10
13	35	100.0	64	2	US-08-765-179B-14
14	35	100.0	70	3	US-08-554-840-9
15	35	100.0	76	4	US-08-851-362D-21
16	35	100.0	79	3	US-08-554-840-14
17	35	100.0	80	3	US-08-554-840-12
18	35	100.0	80	3	US-08-554-840-13
19	35	100.0	80	3	US-08-554-840-15
20	35	100.0	81	3	US-08-554-840-11
21	35	100.0	91	2	US-08-273-146-49
22	35	100.0	93	1	US-08-276-852-111
23	35	100.0	93	1	US-08-899-575-111
24	35	100.0	93	1	US-08-899-575-111
25	35	100.0	93	1	US-08-899-575-111
26	35	100.0	93	1	US-08-899-575-111
27	35	100.0	96	3	US-08-466-368-6

28	35	100.0	103	4	US-09-240-274-42	Sequence 42, Appl
29	35	100.0	104	1	US-08-276-852-92	Sequence 92, Appl
30	35	100.0	104	1	US-08-276-852-94	Sequence 94, Appl
31	35	100.0	104	1	US-08-276-852-100	Sequence 100, Appl
32	35	100.0	104	1	US-08-276-852-106	Sequence 106, Appl
33	35	100.0	104	1	US-08-899-575-92	Sequence 92, Appl
34	35	100.0	104	1	US-08-899-575-94	Sequence 94, Appl
35	35	100.0	104	1	US-08-899-575-100	Sequence 100, Appl
36	35	100.0	104	1	US-08-899-575-106	Sequence 106, Appl
37	35	100.0	104	1	US-08-899-575-92	Sequence 92, Appl
38	35	100.0	104	1	US-08-899-575-94	Sequence 94, Appl
39	35	100.0	104	1	US-08-899-575-100	Sequence 100, Appl
40	35	100.0	104	1	US-08-899-575-106	Sequence 106, Appl
41	35	100.0	104	1	PCT-US95-08743-92	Sequence 92, Appl
42	35	100.0	104	5	PCT-US95-08743-94	Sequence 94, Appl
43	35	100.0	104	5	PCT-US95-08743-100	Sequence 100, Appl
44	35	100.0	104	5	PCT-US95-08743-106	Sequence 106, Appl
45	35	100.0	105	1	US-08-276-852-89	Sequence 89, Appl

#### ALIGNMENTS

```
RESULT 1
PCT-US91-02942-42
Sequence 42 Application PC/PTUS9102942
GENERAL INFORMATION:
APPLICANT: ROTHLEIN, ROBERT
APPLICANT: ADAIR, JOHN R
TITLE OF INVENTION: HUMANIZED CDR-GRAFTED ICAM-1 ANTIBODY
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSER: Sterne, Kessler, Goldstein & Fox
STREET: 1225 Connecticut Ave. NW Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/02942
FILING DATE: 19910429
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9009549.8
FILING DATE: 27-Apr-1990
ATTORNEY/AGENT INFORMATION:
NAME: FOX, SAM L.
REGISTRATION NUMBER: 30,353
REFERENCE/DOCKET NUMBER: 1011.0586600
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 466-0800
TELEFAX: (202) 833-8716
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US91-02942-42
```

Query Match 100.0%; Score 35; DB 5; Length 25;  
Best local similarity 100.0%; Prod. No. 0.68; 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;  
OY 1 TDFTLTI 7  
|||||||

Db 6 TDFTLTI 12

## RESULT 2

US-08-470-139-13  
Sequence 13, Application US/08470139  
Patent No. 5998586  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: Interleukin-5 specific recombinant antibodies  
NUMBER OF SEQUENCES: 28  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,139  
FILING DATE: 06 JUNE-1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: TRUJILLO, DOREEN YATKO  
REGISTRATION NUMBER: 35,719  
REFERENCE/DOCKET NUMBER: CARP-0044  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-470-139-13

Query Match 100.0%; Score 35; DB 2; Length 32;  
Best Local Similarity 100.0%; Pred. No. 0.88; 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFTLTI 7  
|||||  
Db 13 TDFTLTI 19

RESULT 3  
US-08-569-147-40  
Sequence 40, Application US/08569147  
Patent No. 6180377  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: HUMANISED ANTIBODIES  
NUMBER OF SEQUENCES: 95  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &  
ADDRESS: No. 6180377f1s, LLP  
STREET: One Liberty Place - 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: U.S.A.  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.26 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/569,147  
FILING DATE: 25-March-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: TRUJILLO, DOREEN YATKO  
REGISTRATION NUMBER: 35,719  
REFERENCE/DOCKET NUMBER: CARP-0047  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-3100

TELEFAX: (215) 568-3439

INFORMATION FOR SEQ ID NO: 40:

SEQUENCE CHARACTERISTICS:

LENGTH: 32 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-569-147-40

Query Match 100.0%; Score 35; DB 4; Length 32;

Best Local Similarity 100.0%; Pred. No. 0.88; 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFTLTI 7  
|||||  
Db 13 TDFTLTI 19

## RESULT 4

US-09-347-061-13  
Sequence 13, Application US/09347061  
Patent No. 6316227  
GENERAL INFORMATION:  
APPLICANT: Bodmer, Mark  
APPLICANT: Altwal, Diljeet Singh  
TITLE OF INVENTION: Emtage, John Spencer  
FILE REFERENCE: CARP-0071  
CURRENT APPLICATION NUMBER: US/09/347,061  
CURRENT FILING DATE: 1999-07-02  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 13  
LENGTH: 32  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: misc.feature  
OTHER INFORMATION: Consensus  
US-09-347-061-13

Query Match 100.0%; Score 35; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 0.88; 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDFTLTI 7  
|||||  
Db 13 TDFTLTI 19

RESULT 5

US-09-425-638A-92  
Sequence 92, Application US/09425638A  
Patent No. 6342587

GENERAL INFORMATION:

APPLICANT: Carlos F. Barbas III, Christoph Rader, Gerd Ritter, Sydney Welt an

APPLICANT: Lloyd J. Old

TITLE OF INVENTION: A33 ANTIGEN SPECIFIC IMMUNOGLOBULIN PRODUCTS AND USES THE

FILE REFERENCE: LUD 5630

CURRENT APPLICATION NUMBER: US/09/425,638A

CURRENT FILING DATE: 1999-10-22

NUMBER OF SEQ ID NOS: 129

SEQ ID NO 92

LENGTH: 32

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

US-09-425-638A-92



Query Match 100.0%; Score 35; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 0.88;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDETLTI 7  
| | | | |  
Db 13 TDETLTI 19

RESULT 5  
US-09-425-638A-94  
; Sequence 94, Application US/09425638A  
; Patent No. 6342587  
; GENERAL INFORMATION:  
; APPLICANT: Carlos F. Barbas III, Christoph Rader, Gerd Ritter, Sydney Welt and  
; APPLICANT: Lloyd J. Old  
; TITLE OF INVENTION: A33 ANTIGEN SPECIFIC IMMUNOGLOBULIN PRODUCTS AND USES THEREOF  
; FILE REFERENCE: LUD 5630  
; CURRENT APPLICATION NUMBER: US/09/425,638A  
; CURRENT FILING DATE: 1999-10-22  
; NUMBER OF SEQ ID NOS: 129  
; SEQ ID NO 94  
; LENGTH: 32  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-425-638A-94

Query Match 100.0%; Score 35; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 0.88;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDETLTI 7  
| | | | |  
Db 13 TDETLTI 19

RESULT 7  
US-09-543-004-92  
; Sequence 92, Application US/09543004  
; Patent No. 6346249  
; GENERAL INFORMATION:  
; APPLICANT: Carlos F. Barbas III, Christoph Rader, Gerd Ritter, Sydney Welt and  
; APPLICANT: Lloyd J. Old  
; TITLE OF INVENTION: A33 ANTIGEN SPECIFIC IMMUNOGLOBULIN PRODUCTS AND USES THEREOF  
; FILE REFERENCE: LUD 5630.1  
; CURRENT APPLICATION NUMBER: US/09/543,004  
; CURRENT FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: 09/425,638  
; PRIOR FILING DATE: 1999-10-22  
; NUMBER OF SEQ ID NOS: 129  
; SEQ ID NO 92  
; LENGTH: 32  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-543-004-92

Query Match 100.0%; Score 35; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 0.88;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDETLTI 7  
| | | | |  
Db 13 TDETLTI 19

RESULT 8  
US-09-543-004-94  
; Sequence 94, Application US/09543004

; Patent No. 6346249  
; GENERAL INFORMATION:  
; APPLICANT: Carlos F. Barbas III, Christoph Rader, Gerd Ritter, Sydney Welt and  
; APPLICANT: Lloyd J. Old  
; TITLE OF INVENTION: A33 ANTIGEN SPECIFIC IMMUNOGLOBULIN PRODUCTS AND USES THEREOF  
; FILE REFERENCE: LUD 5630.1  
; CURRENT APPLICATION NUMBER: US/09/543,004  
; CURRENT FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: 09/425,638  
; PRIOR FILING DATE: 1999-10-22  
; NUMBER OF SEQ ID NOS: 129  
; SEQ ID NO 94  
; LENGTH: 32  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-543-004-94

Query Match 100.0%; Score 35; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 0.88;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDETLTI 7  
| | | | |  
Db 13 TDETLTI 19

RESULT 9  
PCT-US91-02942-9  
; Sequence 9, Application PC/TUS9102942  
; GENERAL INFORMATION:  
; APPLICANT: ROTHSTEIN, ROBERT  
; APPLICANT: ADAIR, JOHN R  
; APPLICANT: ARIHAL, DILBERT S  
; TITLE OF INVENTION: HUMANIZED CDR-GRAFTED ICAM-1 ANTIBODY  
; NUMBER OF SEQUENCES: 102  
; CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Ave. NW Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036  
; COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/02942  
FILING DATE: 19910429  
CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9009549.8  
FILING DATE: 27-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: FOX, SAM L  
REGISTRATION NUMBER: 30,353  
REFERENCE/DOCKET NUMBER: 1011.0586600  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
; INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
PCT-US91-02942-9

Query Match 100.0%; Score 35; DB 5; Length 50;

Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TDFTLTI 7  
Db 24 TDFTLTI 30

RESULT 10  
US-08-162-102C-43  
; Sequence 43, Application US/08162102C  
; Patent No. 5762905  
; GENERAL INFORMATION:  
; APPLICANT: Burton, Dennis R.  
; APPLICANT: Barbas, III, Carlos F.  
; APPLICANT: Chanock, Robert M.  
; APPLICANT: Murphy, Brian R.  
; APPLICANT: Crowe, Jr., James E.  
; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES  
; TITLE OF INVENTION: TO RESPIRATORY SYNCYTIAL VIRUS  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92037

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/162,102C  
FILING DATE: 10-DEC-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
TELEPHONE: (619) 678-5070  
TELEFAX: (619) 678-5099  
TELECOMMUNICATION INFORMATION:  
REFERENCE/DOCKET NUMBER: 07300/007001  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 52 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-162-102C-43

Query Match 100.0%; Score 35; DB 1; Length 52;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TDFTLTI 7  
Db 13 TDFTLTI 19

RESULT 11  
US-08-162-102C-44  
; Sequence 44, Application US/08162102C  
; Patent No. 5762905  
; GENERAL INFORMATION:  
; APPLICANT: Burton, Dennis R.  
; APPLICANT: Barbas, III, Carlos F.  
; APPLICANT: Chanock, Robert M.  
; APPLICANT: Murphy, Brian R.  
; APPLICANT: Crowe, Jr., James E.

;; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES  
;; TITLE OF INVENTION: TO RESPIRATORY SYNCYTIAL VIRUS  
;; NUMBER OF SEQUENCES: 49  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Fish & Richardson P.C.  
;; STREET: 4225 Executive Square, Suite 1400  
;; CITY: La Jolla  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 92037

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/162,102C  
FILING DATE: 10-DEC-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07300/007001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 678-5070  
TELEFAX: (619) 678-5099  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 53 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-162-102C-44

Query Match 100.0%; Score 35; DB 1; Length 53;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TDFTLTI 7  
Db 13 TDFTLTI 19

RESULT 12  
US-08-765-179B-10  
; Sequence 10, Application US/08765179B  
; Patent No. 5854027  
; GENERAL INFORMATION:  
; APPLICANT: STEINBACHER, Stefan  
; TITLE OF INVENTION: PROCESS FOR MODIFYING THE STABILITY  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nikolaïdo, Marmelstein, Murray & Oram LLP  
; STREET: 655 Fifteenth Street N.W. Suite 330  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-5701

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/765,179B  
FILING DATE: 14-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/02626

FILED DATE: 06-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 25 115.7  
FILING DATE: 15-JUL-1994  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 64 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-765-179B-10

Query Match 100.0%; Score 35; DB 2; Length 64;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
|||||  
DB 38 TDFTLTI 44

RESULT 13  
US-08-765-179B-14  
Sequence 14, Application US/08765179B  
Patent No. 5854027  
GENERAL INFORMATION:  
APPLICANT: STEINBACHER, Boris  
APPLICANT: STEINBACHER, Stefan  
TITLE OF INVENTION: PROCESS FOR MODIFYING THE STABILITY  
TITLE OF INVENTION: OF ANTIBODIES  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nikaïdo, Marmelstein, Murray & Oram LLP  
STREET: 655 Fifteenth Street N.W. Suite 330  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-5701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/765,179B  
FILING DATE: 14-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/02626  
FILING DATE: 06-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 25 115.7  
FILING DATE: 15-JUL-1994  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 64 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-765-179B-14

Query Match 100.0%; Score 35; DB 2; Length 64;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
|||||  
DB 38 TDFTLTI 44

RESULT 14  
US-08-554-840-9  
Sequence 9, Application US/08554840  
Patent No. 6001358  
GENERAL INFORMATION:  
APPLICANT: BLACK, Amelia  
APPLICANT: HANNA, Nabil  
APPLICANT: PADMAN, Eduardo A.  
APPLICANT: NEWMAN, Roland A.  
TITLE OF INVENTION: HUMANIZED ANTIBODIES TO HUMAN gp39,  
COMPOSITIONS CONTAINING AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/554,840  
FILING DATE: 07-NOV-1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-127  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 70 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-554-840-9

Query Match 100.0%; Score 35; DB 3; Length 70;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
|||||  
DB 51 TDFTLTI 57

RESULT 15  
US-08-851-362D-21  
Sequence 21, Application US/08851362D  
Patent No. 6235883  
GENERAL INFORMATION:  
APPLICANT: Jakobovits, Aya  
APPLICANT: Yang, Xiao-Dong  
APPLICANT: Gallo, Michael  
APPLICANT: Jia, Xiao-Chi  
TITLE OF INVENTION: Human Monoclonal Antibodies to Epidermal  
Growth Factor Receptor  
FILE REFERENCE: Cell 4.20  
CURRENT APPLICATION NUMBER: US/08/851,362D  
CURRENT FILING DATE: 1997-05-05  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 21  
LENGTH: 76

;  
; TYPE: PRT  
; ORGANISM: human  
US-08-851-362D-21

Query Match 100.0%; Score 35; DB 4; Length 76;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 TDFTLTI 7  
|||||  
DB 50 TDFTLTI 56

Search completed: July 15, 2002, 12:59:18  
Job time: 389 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:27:08 ; Search time 49.13 Seconds

(without alignments)  
13.691 Million cell updates/sec

Title: US-09-712-819a-5

Perfect score: 35

Sequence: 1 TDFTLTI 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 455

Minimum DB seq length: 0  
Maximum DB seq length: 7

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	37.1	6	PD0028	pev-kinin 2 - pena
2	12	34.3	5	A32516	cholecystokinin-5
3	12	34.3	5	PT0644	T-cell receptor be
4	12	34.3	7	B39127	phosphotransferase
5	12	34.3	7	PT0655	T-cell receptor be
6	11	31.4	4	I40597	biotin A - citroba
7	11	31.4	7	B30608	Ig kappa chain V-1
8	10	28.6	6	PT0630	T-cell receptor be
9	10	28.6	7	S09066	globulin IV alpha
10	9	25.7	3	T13892	cytochrome-c oxida
11	9	25.7	5	E42364	flagellar protein
12	9	25.7	5	S68274	major protein anti
13	9	25.7	5	S68326	blood cell protein
14	9	25.7	5	S69237	surface protein te
15	9	25.7	6	A60986	N-formyl oligopept
16	9	25.7	6	B44510	hypothetical prote
17	9	25.7	6	A43766	28k ubiquitin-limu
18	9	25.7	6	I37263	Y protein - human
19	9	25.7	6	S29881	Na+/K+-exchanging
20	9	25.7	6	I65346	MHC H2-L antigen
21	9	25.7	6	PT0587	T-cell receptor be
22	9	25.7	7	S25266	pile protein - Esc
23	9	25.7	7	A25269	sex pheromone CAM3
24	9	25.7	7	A30812	sex pheromone CCPI
25	9	25.7	7	PT0611	T-cell receptor be
26	8	22.9	4	PT0697	T-cell receptor be
27	8	22.9	5	I39964	ribosomal protein
28	8	22.9	5	I39966	ribosomal protein
29	8	22.9	5	I39965	ribosomal protein

30	8	22.9	5	2	A44692	fulicin - giant Af
31	8	22.9	5	2	PT0729	T-cell receptor be
32	8	22.9	5	2	PT0590	T-cell receptor be
33	8	22.9	5	2	G44817	27.5 kDa structural
34	8	22.9	5	2	I44817	27.5K structural p
35	8	22.9	5	2	E44817	27.5K structural p
36	8	22.9	5	2	C44817	28.5K structural p
37	8	22.9	5	2	A44817	28K structural pro
38	8	22.9	6	2	A19780	transferrin - bovi
39	8	22.9	6	2	A46474	fc epsilon RIIb -
40	8	22.9	6	2	PT0637	T-cell receptor be
41	8	22.9	6	2	PT0641	T-cell receptor be
42	8	22.9	7	2	E61491	seed protein ws-5
43	8	22.9	7	2	PS0254	18K protein 5507 -
44	8	22.9	7	2	PT0642	T-cell receptor be
45	8	22.9	7	2	PT0689	T-cell receptor be

#### ALIGNMENTS

RESULT 1  
PD0028  
pev-kinin 2 - penaeid shrimp (Penaeus vannamei) (fragment)  
C:Species: Penaeus vannamei  
C>Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 19-May-2000  
C:Accession: PD0028  
R:Nieto, J.; Veeleart, D.; Derva, R.; Waelkens, E.; Cerstiaens, A.; Coast, G.; Devree  
Biochem. Biophys. Res. Commun. 246, 406-411, 1998  
A>Title: Identification of one tachykinin- and two kinin-related peptides in the brai  
A:Reference number: PD0027; MUID:98342103  
A:Accession: PD0028  
A:Molecule type: protein  
A:Residues: 1-6 <NTE>  
C:Comment: This peptide belongs to myotropic neuropeptides.

Query Match 37.1% Score 13; DB 2; Length 6;  
Best Local Similarity 66.7%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 DFT 4  
Db 1 DFS 3

RESULT 2  
A32516  
cholecystokinin-5 - dog  
N:Alternate names: CCK-5  
C:Species: Canis lupus familiaris (dog)  
C>Date: 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 18-Aug-2000  
C:Accession: A32516  
R:Shively, J.; Reeve Jr., J.R.; Eysselein, V.E.; Ben-Avram, C.; Vigna, S.R.; Walsh, J  
Am. J. Physiol. 252, G272-G275, 1987  
A>Title: CCK-5: sequence analysis of a small cholecystokinin from canine brain and in  
A:Reference number: A32516; MUID:97153871  
A:Accession: A32516  
A:Molecule type: protein  
A:Residues: 1-5 <SHI>  
C:Comment: This peptide corresponds to the five carboxyl-terminal residues of cholecy  
C:Superfamily: gastrin  
C:Keywords: amidated carboxyl end; neuropeptide  
F:5/Modified site: amidated carboxyl end (Phe) #status experimental

Query Match 34.3% Score 12; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 2 DF 3  
Db 4 DF 5

RESULT 3  
PT0644  
T-cell receptor beta chain V-D-J region (111-16) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
C:Accession: PT0644  
R:Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
A:Reference number: PT0509; MUID:91277601  
A:Accession: PT0644  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-5 <PEP>  
A:Experimental source: newborn thymus, strain BALB/c  
C:Keywords: T-cell receptor

Query Match 34.3%; Score 12; DB 2; Length 5;  
Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 1 TDFT 4  
DB 2 SSFT 5

RESULT 4  
B39127  
phosphotransferase system enzyme II (EC 2.7.1.69) - Escherichia coli (fragment)  
C:Species: Escherichia coli  
C:Date: 27-Nov-1991 #sequence\_revision 27-Nov-1991 #text\_change 08-Oct-1999  
C:Accession: B39127  
R:Hardisty, C.; Ferran, C.; DiRienzo, J.M.  
J. Bacteriol. 173, 449-456, 1991  
A:Title: Plasmid-mediated sucrose metabolism in Escherichia coli: characterization of sc  
rIn.  
A:Reference number: A39127; MUID:91100329  
A:Accession: B39127  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-7 <HAR>  
A:Cross-references: GB:M38416; NID:g155142; PIDN:AAA98418.1; PID:g155144  
C:Keywords: phosphotransferase

Query Match 34.3%; Score 12; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DF 3  
DB 2 DF 3

RESULT 5  
PT0665  
T-cell receptor beta chain V-D-J region (121-38M) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
C:Accession: PT0665  
R:Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
A:Reference number: PT0509; MUID:91277601  
A:Accession: PT0665  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-7 <PEP>  
A:Experimental source: day 4 postnatal thymus, strain BALB/c  
C:Keywords: T-cell receptor

Query Match 34.3%; Score 12; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 DF 3  
DB 6 DF 7

RESULT 6  
I40697  
biotin A - Citrobacter freundii (fragment)  
C:Species: Citrobacter freundii  
C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 12-Aug-1996  
C:Accession: I40697  
R:Shuan, D.; Campbell, A.  
Gene 67, 203-211, 1988  
A:Title: Transcriptional regulation and gene arrangement of Escherichia coli, Citrob  
A:Reference number: I40697; MUID:89006280  
A:Accession: I40697  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-4 <RES>  
A:Cross-references: GB:M21922; NID:g144434

Query Match 31.4%; Score 11; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TD 2  
DB 2 TD 3

RESULT 7  
E30608  
Ig kappa chain V-II region (Gag) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 29-Jun-1989 #sequence\_revision 29-Jun-1989 #text\_change 16-Aug-1996  
C:Accession: E30608  
R:Goni, F.R.; Chen, P.P.; McGinlis, D.; Arjonailla, M.L.; Fernandez, J.; Carson, D.; S  
J. Immunol. 142, 3158-3163, 1989  
A:Title: Structural and idiotypic characterization of the L chains of human IgM auto  
A:Reference number: A30601; MUID:89215279  
A:Accession: E30608  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-7 <GON>  
C:Keywords: heterotetramer; immunoglobulin

Query Match 31.4%; Score 11; DB 2; Length 7;  
Best Local Similarity 40.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 DF 6  
DB 1 ETVT 5

RESULT 8  
PT0650  
T-cell receptor beta chain V-D-J region (121-38F) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
C:Accession: PT0650  
R:Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions  
A:Reference number: PT0509; MUID:91277601

A:Accession: F10650  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-6 <FE>  
A:Experimental source: day 4 postnatal thymus, strain BALB/c  
C:Keywords: T-cell receptor

Query Match 28.6%; Score 10; DB 2; Length 6;  
Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TDFL 4  
1 1  
Db 3 SDAF 6

RESULT 9  
S09066  
globulin IV alpha subunit delta-1 chain, seed - cucurbit (fragments)  
N:Alternate names: 11S globulin alpha subunit delta-1 chain  
C:Species: Cucurbita sp. (cucurbit)  
C:Date: 21-Nov-1993 #sequence\_revision 08-Nov-1996 #text\_change 08-Nov-1996  
C:Accession: S09066  
R:Omura, M.; Hara, I.; Matsubara, H.  
Plant Cell Physiol. 21, 157-167, 1980  
A:Title: Pumpkin (Cucurbita sp.) seed globulin IV. Terminal sequences of the acidic and  
A:Reference number: S09066  
A:Accession: S09066  
A:Molecule type: protein  
A:Residues: 1-6; 7 <OHM>

Query Match 28.6%; Score 10; DB 2; Length 7;  
Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 DFTL 5  
1 1  
Db 3 DFTL 6

RESULT 10  
T13892  
cytochrome-c oxidase (EC 1.9.3.1) chain I [imported] - river lamprey mitochondrion (frag  
C:Species: mitochondrion lampetra fluviatilis (river lamprey)  
C:Date: 15-Jun-2001 #sequence\_revision 15-Jun-2001 #text\_change 15-Jun-2001  
C:Accession: T13892  
R:Delafre, C.; Barriel, V.; Jallier, S.; Janvier, P.; Gachelin, G.  
Mol. Biol. Evol. 14, 807-813, 1997  
A:Title: The main features of the craniate mitochondrial DNA between the ND1 and the COI  
A:Reference number: Z17775; MUID:97398704  
A:Accession: T13892  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3 <DE>  
A:Cross-references: EMBL:Y09528; NID:q2340016; PIDN:CAA70721.1; PID:q4379123  
C:Genetics:  
A:Genome: mitochondrion  
A:Note: COI  
C:Keywords: mitochondrion; oxidoreductase

Query Match 25.7%; Score 9; DB 3; Length 3;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 TL 5  
1 1  
Db 2 TL 3

RESULT 11

E42364  
flagellar protein flir - Salmonella typhimurium (fragment)  
C:Species: Salmonella typhimurium  
C:Date: 24-Jul-1992 #sequence\_revision 24-Jul-1992 #text\_change 30-Sep-1993  
C:Accession: E42364  
R:Voegel, A.P.; Homma, M.; Irkura, V.M.; Macnab, R.M.  
J. Bacteriol. 173, 3564-3572, 1991  
A:Title: Salmonella typhimurium mutants defective in flagellar filament regrowth and  
A:Reference number: A42364; MUID:91258342  
A:Accession: E42364  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-5 <VO>  
A:Cross-references: GH:M62408

Query Match 25.7%; Score 9; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 TL 5  
1 1  
Db 3 TL 4

RESULT 12  
E60274  
major protein antigen MPB63 - Mycobacterium tuberculosis (fragment)  
C:Species: Mycobacterium tuberculosis  
C:Date: 11-Dec-1992 #sequence\_revision 11-Dec-1992 #text\_change 30-Sep-1993  
C:Accession: E60274  
R:Nagal, S.; Wiker, H.G.; Harboe, M.; Kinnomoto, M.  
Infect. Immun. 59, 372-382, 1991  
A:Title: Isolation and partial characterization of major protein antigens in the cult  
A:Reference number: A60274; MUID:91099989  
A:Accession: E60274  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-5 <MG>

Query Match 25.7%; Score 9; DB 2; Length 5;  
Best Local Similarity 25.0%; Pred. No. 2.8e+05;  
Matches 1; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 FTLT 6  
1 1  
Db 2 YFTL 5

RESULT 13  
S68326  
blood cell protein B - Ascidia ceratodes (fragment)  
N:Alternate names: Abcp-B  
C:Species: Ascidia ceratodes  
C:Date: 19-Mar-1997 #sequence\_revision 29-Aug-1997 #text\_change 11-Jun-1999  
C:Accession: S68326  
R:Taylor, S.W.; Ross, M.M.; Waite, J.H.  
Arch. Biochem. Biophys. 324, 228-240, 1995  
A:Title: Novel 3,4-di- and 3,4,5-trihydroxyphenylalanine-containing polypeptides from  
A:Reference number: S68325; MUID:96132650  
A:Accession: S68326  
A:Molecule type: protein  
A:Residues: 1-5 <TA>  
P:2/Modified site: 3',4',5'-trihydroxyphenylalanine (Tyr) #status experimental  
F:4/Modified site: 3',4'-dihydroxyphenylalanine (Tyr) #status experimental

Query Match 25.7%; Score 9; DB 2; Length 5;  
Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 DF 3

Db 1 DY 2

# RESULT 14

S69237  
surface protein tetraabrachion heavy chain - Staphylothermus marinus (Fragment)  
C:Species: Staphylothermus marinus  
C:Date: 04-Dec-1997 #sequence\_revision 04-Dec-1997 #text\_change 17-Mar-1999  
C:Accession: S69237  
R:Peterson, J.; Nitsch, M.; Kuehlmoergen, B.; Goldik, R.; Lupas, A.; Kellermann, J.; Engelth  
J. Mol. Biol. 245: 385-401, 1995  
A:Title: Tetraabrachion: A filamentous Archaeobacterial surface protein assembly of unusual  
A:Reference number: S69237; MUID:95139068  
A:Accession: S69237  
A:Molecule type: protein  
A:Residues: 1-5 <PEP>  
A:Experimental source: strain FI, DSM 3639  
C:Keywords: cell wall; glycoprotein; heat-stable protein

## Query Match

Best Local Similarity 25.7%; Score 9; DB 2; Length 5;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 TL 5  
||  
Db 2 TL 3

# RESULT 15

A60986

N-formyl oligopeptide - Escherichia coli (fragment)  
C:Species: Escherichia coli  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 31-Dec-1993  
C:Accession: A60986  
R:Broom, M.F.; Mellor, D.M.; Chadwick, V.S.  
Experientia 45, 1097-1099, 1989  
A:Title: Purification and amino acid sequencing of naturally occurring N-formyl-methionyl  
A:Reference number: A60986; MUID:90092408  
A:Accession: A60986  
A:Molecule type: protein  
A:Residues: 1-6 <BRO>  
A:Comment: This hexapeptide was the longest of several N-formyl oligopeptides reported.  
F.I/Modified site: N-formylmethionine #status experimental

## Query Match

Best Local Similarity 25.7%; Score 9; DB 2; Length 6;  
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 FTL 5  
|||  
Db 3 FTL 5

Search completed: July 15, 2002, 13:27:09  
Job time: 439 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:58:10 ; Search time 24.49 Seconds  
(without alignments)  
11.067 Million cell updates/sec

Title: US-09-712-819A-5  
Perfect score: 35  
Sequence: 1 TDFTLTI 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 84

Minimum DB seq length: 0  
Maximum DB seq length: 7

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match length	DB ID	Description
1	14	40.0	7 1	AL17_CYDPO P82158 cydia pomon
2	12	34.3	5 1	RE11_LITRU P82070 litorea rub
3	11	31.4	5 1	B10A_CITFR P13071 citrobacter
4	11	31.4	5 1	RE31_LITRU P82072 litorea rub
5	11	31.4	5 1	RE32_LITRU P82073 litorea rub
6	10	28.6	5 1	E104_LITRU P83100 litorea rub
7	9	25.7	7 1	CCP1_ENTFA P20104 enterococu
8	9	25.7	7 1	C1A_ENTFA P11932 enterococu
9	8	22.9	5 1	AL14_CARMA P81817 carcius ma
10	8	22.9	5 1	PSK_DAUCA P58261 daucus caro
11	8	22.9	5 1	RE21_LITRU P82071 litorea rub
12	8	22.9	7 1	AL12_CARMA P81805 carcius ma
13	8	22.9	7 1	AL13_CARMA P81806 carcius ma
14	8	22.9	7 1	AL14_CARMA P81807 carcius ma
15	8	22.9	7 1	AL15_CARMA P81808 carcius ma
16	8	22.9	7 1	FAR1_ASCSU P31889 ascariis suu
17	8	22.9	4 1	RM01_YESTY P36513 saccharomyc
18	7	20.0	6 1	FAR1_MONEX P41966 monelia ex
19	7	20.0	6 1	LOK1_LOCHI P41491 locusta mi9
20	7	20.0	6 1	UN06_CLOPA P81351 clostridium
21	7	20.0	7 1	FAR1_PROCL P38499 procambarus
22	7	20.0	7 1	FAR1_PANRE P43498 panagrellus
23	7	20.0	7 1	FAR1_PANRE P41865 calliphora
24	7	20.0	7 1	FAR1_CALYO P99025 mus musculu
25	7	20.0	7 1	GRP_MOUSE P06294 dactylinum d
26	7	20.0	7 1	GAO_DACDE P24272 vibrio lisc
27	6	17.1	3 1	LOXE_VIBFI P35904 acinetiba fu
28	6	17.1	4 1	ACH1_ACHFU P42562 hirudo medi
29	6	17.1	4 1	FAR3_HIRME P42563 hirudo medi
30	6	17.1	4 1	FAR4_HIRME P58705 anophleura
31	6	17.1	4 1	PEKA_HIRME P42561 hirudo medi
32	6	17.1	4 1	PEKA_HIRME P58707 anophleura
33	6	17.1	4 1	ELRN_ANTEL

34	6	17.1	4 1	FMRF_MACNI P01162 macrocallis
35	6	17.1	4 1	FYRI_ANTEL P58706 anophleura
36	6	17.1	4 1	OCP1_OCTMI P58648 octopus min
37	6	17.1	4 1	OCP3_OCTMI P58649 octopus min
38	6	17.1	5 1	ER03_LITRU P82099 litorea rub
39	6	17.1	5 1	FARP_ARTTR P41853 arthroposibi
40	6	17.1	5 1	FAP2_PARMA P81864 pardechirus
41	6	17.1	5 1	SOGA_ACHDO P19991 acheta dome
42	6	17.1	5 1	TPIS_CANFA P54714 canis famli
43	6	17.1	5 1	TRM3_ECOLI P13973 escherichia
44	6	17.1	5 1	UC22_MAIZE P80628 zea mays (m
45	6	17.1	5 1	UXA4_CHLTR P38005 chlamydia t

## ALIGNMENTS

RESULT 1				
ALL7_CYDPO	STANDARD:	PRT:	7 AA.	
ID ALL7_CYDPO				
AC P82158:	30-MAY-2000 (Rel. 39, Created)			
DT 30-MAY-2000	(Rel. 39, Last sequence update)			
DT 30-MAY-2000	(Rel. 39, Last annotation update)			
DE Cydiastatin 7.				
OS Cydia pomonella (Coddling moth).				
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditytysia;				
OC Tortricidae; Tortricidae; Olethreutinae; Cydia.				
OX NCBI_TaxID=82600;				
RN [1]				
RP TISSUE=Larva;				
RC MEDLINE=98054539; PubMed=9392879;				
RX Dove H., Johnson A.H., Maestro J.-L., Scott A.G., Winstanley D.,				
RA Davey M., East P.D., Thorpe A.;				
RT "Lepidopteran peptides of the allatostatin superfamily.";				
RT Peptides 18:1301-1309(1997).				
CC -! SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.				
KW Neuropeptide; Amidation.				
KM MOD_RES 7 AA: 873 MW: 672879CABBS69350 CRC64:				
FT SEQUENCE				

Query Match 40.0%; Score 14; DB 1; Length 7;  
Best Local Similarity 75.0%; Pred. No. 1e+05;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 DFTL 5  
1 1 1  
DB 4 DFTL 7

RESULT 2  
ID RE11\_LITRU STANDARD: PRT: 5 AA.  
AC P82070:  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Rubellidin 1.1.  
OS Litorea rubella (Desert tree frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litorea.  
OX NCBI\_TaxID=104895;  
RN [1]  
RP SEQUENCE, AND MASS SPECTROMETRY.  
RC TISSUE=Skin secretion;  
RA Steinboerner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,  
RA Tyler M.J., Wallace J.C.;

\*The structure of new peptides from the Australian red tree frog  
'Litorea rubella', the skin peptide profile as a probe for the study

RT of evolutionary trends of amphibians.",  
 RU Aust. J. Chem. 49:955-963(1996).  
 CC -1- FUNCTION: SHOW NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC  
 CC ACTIVITY.  
 CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
 CC -1- MASS SPECTROMETRY: MW=598; METHOD=FAH.  
 KW Amphibian skin.  
 SO SEQUENCE 5 AA: 598 MW: 6DD9C9CAB2A00000 CRC64:

Query Match 34.3%; Score 12; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DF 3  
 DB 2 DF 3

## RESULT 3

BIO\_CITR STANDARD: PRT; 5 AA.  
 ID BIO\_CITR  
 AC P13071;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Adenosylmethionine-8-amino-7-oxononanoate aminotransferase  
 DE (EC 2.6.1.62) (7,8-diamino-pelargonic acid aminotransferase) (DAPA  
 DE aminotransferase) (Fragment).  
 GN BIO.  
 OS Citrobacter freundii.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Citrobacter.  
 OX NCBI\_Taxid=546;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=69006280; PubMed=2971595;  
 RA Shivan D., Campbell A.;  
 RT "Transcriptional regulation and gene arrangement of Escherichia coli,  
 RT Citrobacter freundii and Salmonella typhimurium biotin operons.";  
 RL Gene 67:203-211(1988).  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + 8-amino-7-oxononanoate -> S-adenosyl-4-methylthio-2-oxobutanoate + 7,8-diaminononanoate.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -1- PATHWAY: BIOTIN BIOSYNTHESIS.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-slb.ch/announce/> or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).  
 CC -----  
 CC EMBL: M21923; -; NOT ANNOTATED\_CDS.  
 DR InterPro: IPR000554; AminoTran\_3.  
 DR PROSITE: PS00600; AA\_TRANSFER\_CLASS\_3; PARTIAL.  
 KW Biotin biosynthesis; Transferase; Amino transferase;  
 KW Pyridoxal phosphate.  
 FT NON\_TER  
 SO SEQUENCE 5 AA: 582 MW: 6AAAB1B1A6F00000 CRC64:

Query Match 31.4%; Score 11; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TD 2

DB 3 TD 4

## RESULT 4

ID RE31\_LITRU STANDARD: PRT; 5 AA.  
 AC P82072;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Rubellidin 3.1.  
 OS Litoria rubella (Desert tree frog).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
 OC Litoria.  
 OX NCBI\_Taxid=104895;  
 RN [1]  
 RP SEQUENCE, AND MASS SPECTROMETRY.  
 RC TISSUE-Skin secretion;  
 RA Steinborner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,  
 RA Tyler M.J., Wallace J.C.;  
 RT "The structure of new peptides from the Australian red tree frog 'Litoria rubella', the skin peptide profile as a probe for the study of evolutionary trends of amphibians.";  
 RL Aust. J. Chem. 49:955-963(1996).  
 CC -1- FUNCTION: SHOW NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC ACTIVITY.  
 CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
 CC -1- MASS SPECTROMETRY: MW=655; METHOD=FAH.  
 KW Amphibian skin; Amidation.  
 FT MOD\_RES  
 SO SEQUENCE 5 AA: 656 MW: 71A9C9CB10300000 CRC64:

Query Match 31.4%; Score 11; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 FT 4  
 DB 4 FT 5

## RESULT 5

ID RE32\_LITRU STANDARD: PRT; 5 AA.  
 AC P82073;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Rubellidin 3.2.  
 OS Litoria rubella (Desert tree frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
 OC Litoria.  
 OX NCBI\_Taxid=104895;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE-Skin secretion;  
 RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;  
 RT "Peptides from the skin glands of the Australian buzzing tree frog Litoria electrica. Comparison with the skin peptides from Litoria rubella.";  
 RL Aust. J. Chem. 52:0-0(1999).  
 CC -1- FUNCTION: SHOW NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC ACTIVITY.  
 CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
 KW Amphibian skin.  
 SO SEQUENCE 5 AA: 570 MW: 71A9C9CB62A00000 CRC64:

Query Match 31.4%; Score 11; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 FT 4  
| |  
Db 4 FT 5

RESULT 6  
E104\_LITRU STANDARD; PRT; 5 AA.  
AC P82100;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DE Electrin 4.  
OS Litoria rubella (Desert tree frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litoria.  
OX NCBI\_TaxID=104895;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Skin secretion;  
RA Webnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;  
RT "Peptides from the skin glands of the Australian buzzing tree frog  
Litoria electrica. Comparison with the skin peptides from Litoria  
rubella".  
RL Aust. J. Chem. 52:0-0(1999).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
KW Amphibian skin; Amidation.  
FT MOD\_RES 5  
SO SEQUENCE 5 AA; 616 MW; 61F2D1A059A00000 CRC64;

Query Match 28.6%; Score 10; DB 1; Length 5;  
Best Local Similarity 33.3%; Pred. No. 1e+05; Indels 0; Gaps 0;  
Matches 1; Conservative 2; Mismatches 0;

OY 5 LTI 7  
| |  
Db 2 ITV 4

RESULT 7  
CCFL\_ENTFA STANDARD; PRT; 7 AA.  
AC P20104;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-FEB-1991 (Rel. 17, Last annotation update)  
DE Sex pheromone CCF10.  
OS Enterococcus faecalis (Streptococcus faecalis).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;  
OC Enterococcus.  
OX NCBI\_TaxID=1351;  
RN [1]  
RP SEQUENCE.  
RA MEDLINE=89008313; PubMed=3139658;  
RA Mori M., Sakagami Y., Ishii Y., Isogai A., Kitada C., Fujino M.,  
RA Asagi J.C., Dunny G.M., Suzuki A.;  
RT "Structure of ccf10, a peptide sex pheromone which induces  
RT conjugative transfer of the Streptococcus faecalis tetracycline  
RT resistance plasmid pCF10."  
RL J. Biol. Chem. 263:14574-14578(1988).  
CC -1- FUNCTION: CCF10 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE  
CC HEMOLYSIN PLASMID PCF10.  
DR PIR; A30812; A30812.  
KW Pheromone.  
SO SEQUENCE 7 AA; 790 MW; 72C9D2C731B2C740 CRC64;

Query Match 25.7%; Score 9; DB 1; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 TIL 5  
| |  
Db 3 TIL 4

RESULT 8  
CIA\_ENTFA STANDARD; PRT; 7 AA.  
AC P11932;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 01-FEB-1991 (Rel. 17, Last annotation update)  
DE Sex pheromone CAM373 (Clumping-inducing agent) (CIA).  
OS Enterococcus faecalis (Streptococcus faecalis).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;  
OC Enterococcus.  
OX NCBI\_TaxID=1351;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=87005252; PubMed=309376;  
RA Mori M., Tanaka H., Sakagami Y., Isogai A., Fujino M., Kitada C.,  
RA White B.A., An F.Y., Clewell D.B., Suzuki A.;  
RT "Isolation and structure of the Streptococcus faecalis sex pheromone,  
RT CAM373".  
RL FEBS Lett. 206:69-72(1986).  
CC -1- FUNCTION: CAM373 INDUCES MATING RESPONSE OF DONOR CELLS  
CC HARBORING PAM373.  
CC -1- MISCELLANEOUS: THE N-TERMINUS IS POSSIBLY RESPONSIBLE FOR  
CC SPECIFICITY OF PHEROMONES TO PLASMIDS.  
CC -1- SIMILARITY: C-TERMINAL TO THE PHEROMONES CPD1 AND CAD1.  
DR PIR; A25269; A25269.  
KW Pheromone.  
SO SEQUENCE 7 AA; 734 MW; 75BD072059C05D80 CRC64;

Query Match 25.7%; Score 9; DB 1; Length 7;  
Best Local Similarity 66.7%; Pred. No. 1e+05; Indels 0; Gaps 0;  
Matches 2; Conservative 0; Mismatches 1;

OY 3 FIL 5  
| |  
Db 3 FIL 5

RESULT 9  
AL14\_CARMA STANDARD; PRT; 5 AA.  
ID AL14\_CARMA  
AC P81817;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Carcinustatin 14.  
OS Carcinus maenas (Common shore crab) (Green crab).  
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
OC Eukaryota; Metazoa; Arthropoda; Decapoda; Pleocyemata; Brachyura;  
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
OX NCBI\_TaxID=6759;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
RX MEDLINE=98121193; PubMed=9461295;  
RA Dure H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
RA Thorpe A.;  
RT "Isolation and identification of multiple neuropeptides of the  
RT allatostatin superfamily in the shore crab Carcinus maenas".  
RL Eur. J. Biochem. 250:727-734(1997).  
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
KW Neuropeptide; Amidation; Multigene family.  
FT MOD\_RES 5  
SO SEQUENCE 5 AA; 586 MW; 672879D5A8300000 CRC64;

Query Match 22.9% Score 8; DB 1; Length 5;  
 Best Local Similarity 66.7% Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 FTL 5  
 Db 3 FGL 5

RESULT 10  
 PSK\_DAUCA STANDARD; PRT; 5 AA.  
 AC P58261;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Physulfoalkine-alpha (PSK-alpha) [Contains: Physulfoalkine-beta (PSK-beta)].  
 OS Daucus carota (Carrot).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.  
 OX NCBI\_TaxID=4039;  
 RN [1]  
 RP SEQUENCE, AND IDENTIFICATION BY MASS SPECTROMETRY.  
 RC STRAIN-CV, US-Harumakigusun;  
 RX MEDLINE=20212743; PubMed=10750705;  
 RA Hamai H., Matsuno T., Yamamoto M., Matsubayashi Y., Kobayashi T.,  
 RA Kamada H., Sakagami Y.,  
 RT "A secreted peptide growth factor, physulfoalkine, acting as a  
 stimulatory factor of carrot somatic embryo formation.";  
 RL Plant Cell Physiol. 41:27-32(2000).  
 CC -1- FUNCTION: IN PRESENCE OF 2,4-D, STIMULATES PROLIFERATION OF THE  
 CELLS, BUT DOES NOT STIMULATE DIFFERENTIATION INTO THE SOMATIC  
 EMBRYOS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- PUT: SULFATION IS IMPORTANT FOR ACTIVITY AND FOR THE BINDING TO A  
 CC -1- PUTATIVE MEMBRANE RECEPTOR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE PHYTOSULFOKINE FAMILY.  
 KM Growth factor; Sulfation.  
 FT PEPTIDE 1 4 PHYTOSULFOKINE-BETA.  
 FT MOD\_RES 1 1 SULFATION.  
 FT MOD\_RES 3 3 SULFATION.  
 SQ SEQUENCE 5 AA: 687 MW: 76C1B8504B300000 CRC64;

Query Match 22.9% Score 8; DB 1; Length 5;  
 Best Local Similarity 50.0% Pred. No. 1e+05;  
 Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 FTL 4  
 Db 3 YTL 4

RESULT 11  
 RE21\_LITRU STANDARD; PRT; 5 AA.  
 ID RE21\_LITRU  
 AC P82071;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Rubellidin 2.1  
 OS Litoria rubella (Desert tree frog).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
 CC Litoria.  
 OX NCBI\_TaxID=104895;  
 RN [1]  
 RP SEQUENCE, AND MASS SPECTROMETRY.  
 RC TISSUE=Skin secretion;

RA Steinhorn S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,  
 RA Tyler M.J., Wallace J.C.;  
 RT "The structure of new peptides from the Australian red tree frog  
 RT 'Litoria rubella', the skin peptide profile as a probe for the study  
 RT of evolutionary trends of amphibians.";  
 RL Aust. J. Chem. 49:955-963(1996).  
 CC -1- FUNCTION: SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC  
 CC ACTIVITY.  
 CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
 CC -1- MASS SPECTROMETRY: MW=626; METHOD=FAE.  
 KW Amphibian skin.  
 SQ SEQUENCE 5 AA: 626 MW: 6DD9C9C810300000 CRC64;

Query Match 22.9% Score 8; DB 1; Length 5;  
 Best Local Similarity 50.0% Pred. No. 1e+05;  
 Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 DP 3  
 Db 2 EF 3

RESULT 12  
 ALI2\_CARMA STANDARD; PRT; 7 AA.  
 AC P81605;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Carcinustatin 2.  
 OS Carcinus maenas (Common shore crab) (Green crab).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
 OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
 OX NCBI\_TaxID=6759;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
 RX MEDLINE=98121193; PubMed=9461295;  
 RA Duye H., Johnsen A.H., Maestro J.-L., Scott A.G., Jares P.P.,  
 RA Thorpe A.;  
 RT "Isolation and identification of multiple neuropeptides of the  
 RT allatostatin superfamily in the shore crab Carcinus maenas.";  
 RL Eur. J. Biochem. 250:727-734(1997).  
 CC -1- FUNCTION: MAY ACT AS A NEUROTANSMITTER OR NEUROMODULATOR.  
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 CC Neuropeptide; Amidation; Multigene family.  
 FT MOD\_RES 7 7 AMIDATION (POTENTIAL).  
 SQ SEQUENCE 7 AA: 770 MW: 672879CDBC5DBD70 CRC64;

Query Match 22.9% Score 8; DB 1; Length 7;  
 Best Local Similarity 66.7% Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 FTL 5  
 Db 5 FGL 7

RESULT 13  
 ALI3\_CARMA STANDARD; PRT; 7 AA.  
 ID ALI3\_CARMA  
 AC P81806;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Carcinustatin 3.  
 OS Carcinus maenas (Common shore crab) (Green crab).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
 OC Eubrachyura; Portunoidae; Portunidae; Carcinus.

OX NCBI\_TaxID=6759;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
 RX MEDLINE=98121193; PubMed=9461295;  
 RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
 Thorpe A.;  
 RT "Isolation and identification of multiple neuropeptides of the  
 allatostatin superfamily in the shore crab Carcinus maenas.";  
 RL Eur. J. Biochem. 250:727-734(1997).  
 CC -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
 CC -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 KW Neuropeptide; Multigene family.  
 SO SEQUENCE 7 AA; 796 MW; 672879CDBC8476B70 CRC64;

Query Match 22.9%; Score 8; DB 1; Length 7;  
 Best Local Similarity 66.7%; Pred. No. 1e+05; 1; Indels 0; Gaps 0;  
 Matches 2; Conservative 0; Mismatches 1;

QY 3 FTL 5  
 | |  
 Db 5 FGL 7

RESULT 14  
 ID ALL4\_CARMA STANDARD; PRT; 7 AA.  
 AC P81807;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Carcinus statin 4.  
 OS Carcinus maenas (Common shore crab) (Green crab).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
 OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
 OX NCBI\_TaxID=6759;  
 [1]  
 RN SEQUENCE.  
 RP TISSUE=Cerebral ganglion, and Thoracic ganglion;  
 RC MEDLINE=98121193; PubMed=9461295;  
 RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
 Thorpe A.;  
 RT "Isolation and identification of multiple neuropeptides of the  
 allatostatin superfamily in the shore crab Carcinus maenas.";  
 RL Eur. J. Biochem. 250:727-734(1997).  
 CC -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
 CC -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 KW Neuropeptide; Multigene family.  
 SO SEQUENCE 7 AA; 782 MW; 672879CDBC8476AC0 CRC64;

Query Match 22.9%; Score 8; DB 1; Length 7;  
 Best Local Similarity 66.7%; Pred. No. 1e+05; 1; Indels 0; Gaps 0;  
 Matches 2; Conservative 0; Mismatches 1;

QY 3 FTL 5  
 | |  
 Db 5 FGL 7

RESULT 15  
 ID ALL5\_CARMA STANDARD; PRT; 7 AA.  
 AC P81808;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Carcinus statin 5.  
 OS Carcinus maenas (Common shore crab) (Green crab).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;

OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
 OX NCBI\_TaxID=6759;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
 RX MEDLINE=98121193; PubMed=9461295;  
 RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
 Thorpe A.;  
 RT "Isolation and identification of multiple neuropeptides of the  
 allatostatin superfamily in the shore crab Carcinus maenas.";  
 RL Eur. J. Biochem. 250:727-734(1997).  
 CC -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
 CC -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 KW Neuropeptide; Multigene family.  
 FT MOD.RES 7  
 FT AMIDATION.  
 SO SEQUENCE 7 AA; 781 MW; 672879CDBC8476420 CRC64;

Query Match 22.9%; Score 8; DB 1; Length 7;  
 Best Local Similarity 66.7%; Pred. No. 1e+05; 1; Indels 0; Gaps 0;  
 Matches 2; Conservative 0; Mismatches 1;

QY 3 FTL 5  
 | |  
 Db 5 FGL 7

Search completed: July 15, 2002, 13:38:10  
 Job time: 710 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:37:39 ; Search time 86.73 seconds  
(without alignments)  
13.962 Million cell updates/sec

Title: US-09-712-819A-5  
Sequence: 1 TDFTLVFI 7  
Perfect score: 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 65

Minimum DB seq length: 0  
Maximum DB seq length: 7

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database :  
1: SPREMBL\_19:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.minc:\*  
9: sp.organelle:\*  
10: sp.phage:\*  
11: sp.plant:\*  
12: sp.podent:\*  
13: sp.virus:\*  
14: sp.vertebrate:\*  
15: sp.unclassified:\*  
16: sp.virus:\*  
17: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12	34.3	5	13 P82070	P82070 litorea rub
2	11	31.4	5	13 P82072	P82072 litorea rub
3	11	31.4	5	13 P82073	P82073 litorea rub
4	11	31.4	7	2 Q54248	Q54248 streptomyc
5	10	28.6	5	13 P82100	P82100 litorea rub
6	10	28.6	7	2 Q34028	Q34028 sphingomona
7	9	25.7	7	2 Q07354	Q07354 synechococc
8	9	25.7	7	10 P82445	P82445 nicotiana t
9	8	22.9	5	13 P82071	P82071 litorea rub
10	8	22.9	7	2 P70804	P70804 azotobacter
11	7	20.0	5	2 P83073	P83073 bacillus ce
12	7	20.0	7	2 Q47029	Q47029 enterobacte
13	7	20.0	7	2 P72081	P72081 nocardia la
14	7	20.0	7	8 Q98866	Q98866 spinacia ol
15	6	17.1	5	13 P82099	P82099 litorea rub
16	6	17.1	6	10 P82541	P82541 spinacia ol

17	6	17.1	6	13 P82096	P82096 litorea rub
18	6	17.1	7	2 Q50556	Q50556 actinobacil
19	6	17.1	7	4 Q15903	Q15903 homo sapien
20	6	17.1	7	6 Q28742	Q28742 oryctolagus
21	6	17.1	7	10 Q49223	Q49223 glycine max
22	6	17.1	7	10 Q9C583	Q9C583 arabidopsis
23	6	17.1	7	11 Q63480	Q63480 rattus norv
24	6	17.1	7	11 Q55184	Q55184 rattus norv
25	6	17.1	7	12 Q9Y010	Q9Y010 transmissib
26	6	17.1	7	13 P82065	P82065 litorea rub
27	5	14.3	6	10 P82181	P82181 spinacia ol
28	5	14.3	6	10 P82182	P82182 spinacia ol
29	5	14.3	7	2 Q47505	Q47505 escherichia
30	5	14.3	7	8 Q95945	Q95945 saccharomyc
31	5	14.3	7	10 P93233	P93233 lycopersico
32	5	14.3	7	12 Q67113	Q67113 influenza a
33	5	14.3	7	12 Q65578	Q65578 bovine herp
34	5	14.3	7	13 Q42564	Q42564 fugu rubrip
35	5	14.3	7	15 Q07624	Q07624 rous sarcom
36	4	11.4	4	11 Q08433	Q08433 rattus norv
37	4	11.4	7	4 Q15897	Q15897 homo sapien
38	4	11.4	7	8 P92421	P92421 psathyrosta
39	4	11.4	7	8 P92385	P92385 notoderm mar
40	4	11.4	7	8 P92372	P92372 haynaldia v
41	4	11.4	7	8 P92403	P92403 lophopyrum
42	4	11.4	7	8 P92425	P92425 pseudoroegn
43	4	11.4	7	8 P92387	P92387 henrardia p
44	4	11.4	7	8 P92427	P92427 peridictyon
45	4	11.4	7	8 P92390	P92390 heteranthe

## ALIGNMENTS

RESULT 1  
ID P82070 PRELIMINARY; PRT; 5 AA.  
AC P82070:  
DI 01-MAY-2000 (TREMUREL. 13, Created)  
DT 01-MAY-2000 (TREMUREL. 13, Last sequence update)  
DF 01-MAY-2000 (TREMUREL. 13, Last annotation update)  
DE RUBEALIDIN 1.1.  
OS Litorea rubella (Desert tree frog);  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litorea.  
OX NCBI\_TaxID=104895;  
RN [1]  
RP SEQUENCE, AND MASS SPECTROMETRY.  
RC TISSUE=SKIN SECRETION.  
RA Steinboerner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,  
RA Tyler M.J., Wallace J.C.;  
RT "The structure of new peptides from the Australian red tree frog  
Litorea rubella", the skin peptide profile as a probe for the study  
of evolutionary trends of amphibians.";  
RT Aust. J. Chem. 49:955-963(1996).  
RL -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR  
-1- ANTI-BIOTIC ACTIVITY.  
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
CC -1- MASS SPECTROMETRY: MW=598; METHOD=FAH.  
KW Amphibian skin.  
SQ SEQUENCE 5 AA; 598 MW; 6DD9C9CAB2A00000 CRC64;

Query Match 34.3%; Score 12; DB 13; Length 5;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DF 3  
11  
Db 2 DF 3

```

RESULT 2
ID P82072 PRELIMINARY: PRT: 5 AA.
AC P82072;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
OS RUBELLIDIN 3.1.
OC Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Litoria.
OX NCBI_Taxid=104895;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE=SKIN SECRETION;
RA Steinbörner S.T., Wabnitz P.A., Maugh R.J., Bowie J.H., Gao C.,
RT Tyler M.J., Wallace J.C.;
RT "The structure of new peptides from the Australian red tree frog
RT 'Litoria rubella', the skin peptide profile as a probe for the study
RT of evolutionary trends of amphibians.";
RL Aust. J. Chem. 49:955-963(1996).
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR
CC ANTI-BIOTIC ACTIVITY.
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.
CC KW Amphibian skin; Amidation.
FT MOD_RES 5
FT SEQUENCE 5 AA: 656 MW: 71A9C9CB10300000 CRC64;

```

```

Query Match 31.4%; Score 11; DB 13; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 3 FT 4
DB 4 FT 5

RESULT 3
ID P82073 PRELIMINARY: PRT: 5 AA.
AC P82073;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
OS RUBELLIDIN 3.2.
OC Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Litoria.
OX NCBI_Taxid=104895;
RN [1]
RP SEQUENCE.
RC TISSUE=SKIN SECRETION;
RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;
RT "Peptides from the skin glands of the Australian buzzing tree frog
RT Litoria electrica. Comparison with the skin peptides from Litoria
RT rubella.";
RL Aust. J. Chem. 52:0-0(1999).
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR
CC ANTI-BIOTIC ACTIVITY.
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.
CC KW Amphibian skin.
FT MOD_RES 5
FT SEQUENCE 5 AA: 570 MW: 71A9C9CB62A00000 CRC64;

```

```

Query Match 31.4%; Score 11; DB 13; Length 5;
Best Local Similarity 100.0%; Pred. No. 5.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 FT 4

```

```

DB 4 FT 5

RESULT 4
ID 054248 PRELIMINARY: PRT: 7 AA.
AC 054248;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RPLO PROTEIN (FRAGMENT).
GN RPLO.
OS Streptomyces griseus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomycetes.
OX NCBI_Taxid=1911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2-3-11;
RX MEDLINE=20011291; PubMed=10542330;
RA Poehling S., Plepersberg W., Wehmeler U.F.;
RT "Analysis and regulation of the sec Y gene from Streptomyces griseus
RT N2-3-11 and interaction of the Sec Y protein with the SecA protein.";
RL Biochim. Biophys. Acta 1447:298-302(1999).
DR EMBL: X95915; GAA65160.1; -.
FT NON_TER 1
FT SEQUENCE 7 AA: 760 MW: 72C72B01B2D1B2A0 CRC64;

```

```

Query Match 31.4%; Score 11; DB 2; Length 7;
Best Local Similarity 66.7%; Pred. No. 5.6e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 4 TLT 6
DB 2 TLT 4

RESULT 5
ID P82100 PRELIMINARY: PRT: 5 AA.
AC P82100;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE ELECTRIN 4.
OS Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Litoria.
OX NCBI_Taxid=104895;
RN [1]
RP SEQUENCE.
RC TISSUE=SKIN SECRETION;
RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;
RT "Peptides from the skin glands of the Australian buzzing tree frog
RT Litoria electrica. Comparison with the skin peptides from Litoria
RT rubella.";
RL Aust. J. Chem. 52:0-0(1999).
CC KW Amphibian skin; Amidation.
FT MOD_RES 5
FT SEQUENCE 5 AA: 616 MW: 61F2D1A059A00000 CRC64;

```

```

Query Match 28.6%; Score 10; DB 13; Length 5;
Best Local Similarity 33.3%; Pred. No. 5.6e+05;
Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 5 LTI 7
DB 2 LTI 4

```



```

RESULT 6
ID 034028 PRELIMINARY; PRT; 7 AA.
AC 034028;
DT 01-JUN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CATECHOL-2,3-DIOXYGENASE (FRAGMENT).
GN PHE.
OS Sphingomonas chungbukensis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Sphingomonas.
OX NCBI_TaxID=56193;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN-D077;
RA Kim Y.-C.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88298; AAB6311.1; -.
KM Dioxxygenase.
FT NON_TER
SQ SEQUENCE 7 AA; 868 MW; 71A452D1A699D460 CRC64;

```

```

Query Match 28.6%; Score 10; DB 2; Length 7;
Best Local Similarity 33.3%; Pred. No. 5.6e+05;
Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 5 LTI 7
   : :
DB 3 MTV 5

```

```

RESULT 7
ID 007354 PRELIMINARY; PRT; 7 AA.
AC 007354;
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE NIPK (FRAGMENT).
GN NIPK.
OS Synechococcus sp. (strain PCC 8801 / RF-1) (Cyanothecae PCC 8801).
OC Bacteria; Cyanobacteria; Chroococcales; Cyanothecae.
OX NCBI_TaxID=41431;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN-RF-1;
RX MEDLINE=99231861; Pubmed=10217509;
RA Huang T.C., Lin R.F., Chu M.K., Chen H.M.;
RT "Organization and expression of nitrogen-fixation genes in the aerobic
RT nitrogen-fixing unicellular cyanobacterium Synechococcus sp. strain
RT RF-1."
RL Microbiology 145:743-753(1999).
DR EMBL; AF003700; AAC35193.1; -.
FT NON_TER
SQ SEQUENCE 7 AA; 849 MW; 7412C72AA9D5B030 CRC64;

```

```

Query Match 25.7%; Score 9; DB 2; Length 7;
Best Local Similarity 66.7%; Pred. No. 5.6e+05;
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 3 FTL 5
   : :
DB 3 FDL 5

```

```

RESULT 8
ID P82445 PRELIMINARY; PRT; 7 AA.
AC P82445;

```

```

DT 01-JUN-2000 (TREMblrel. 14, Created)
DT 01-JUN-2000 (TREMblrel. 14, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE 10 KDA CELL WALL PROTEIN (FRAGMENT).
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eunasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1];
RP SEQUENCE.
RC STRAIN=CV. PETIT HAVANA;
RA Blee K.A., Bonham V.A., Mitchell G.P., Robertson D., Slabas A.R.,
RA Wojtaszek P., Bolwell G.P.;
RT "Proteomic study of secondary cell wall proteins from transformed
RT tobacco culture."
RL Planta 0:0-0(2000).
CC -1- SUBCELLULAR LOCATION: CELL WALL.
CC -1- TISSUE SPECIFICITY: XYLEM.
KM Cell wall.
FT NON_TER
SQ SEQUENCE 7 AA; 758 MW; 69D2C1E862D1B2A0 CRC64;

```

```

Query Match 25.7%; Score 9; DB 10; Length 7;
Best Local Similarity 33.3%; Pred. No. 5.6e+05;
Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 5 LTI 7
   : :
DB 1 VTY 3

```

```

RESULT 9
ID P82071 PRELIMINARY; PRT; 5 AA.
AC P82071;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE RUBELLIDIN 2.1.
OS Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Litoria.
OX NCBI_TaxID=104895;
RN [1];
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE=SKIN SECRETION;
RA Steinborner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,
RA Tyler M.J., Wallace J.C.;
RT "The structure of new peptides from the Australian red tree frog
RT 'Litoria rubella', the skin peptide profile as a probe for the study
RT of evolutionary trends of amphibians."
RL Aust. J. Chem. 49:955-963(1996).
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR
CC ANTIHYPERTENSIVE ACTIVITY.
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.
CC -1- MASS SPECTROMETRY: MW=626; METHOD=FA-B.
KM Amphibian skin.
SQ SEQUENCE 5 AA; 626 MW; 6DD9C9CB10300000 CRC64;

```

```

Query Match 22.9%; Score 8; DB 13; Length 5;
Best Local Similarity 50.0%; Pred. No. 5.6e+05;
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 DF 3
   : :
DB 2 EF 3

```

```

RESULT 10

```

P70804  
ID P70804 PRELIMINARY; PRT; 7 AA.  
AC P70804;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE ALGT PROTEIN (FRAGMENT).  
GN ALGT.  
OS Azotobacter vinelandii.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Azotobacter.  
OX NCBI\_TaxID=354;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=;  
RX MEDLINE=96427318; PubMed=8830682;  
RA Rehm B.H.A., Ertesvag H., Valla S.;  
RT "A new Azotobacter vinelandii manuronan C-5-epimerase gene (algG) is  
RT part of an alg gene cluster physically organized in a manner similar  
RT to that in Pseudomonas aeruginosa.";  
RL J. Bacteriol. 178:5884-5889(1996).  
DR EMBL: X87973; CAA61230.1; -.  
FT NON\_TER  
SQ SEQUENCE 7 AA; 684 MW; 71BSA5A5A2D1AED0 CRC64;

Query Match 22.9%; Score 8; DB 2; Length 7;  
Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 TT 7  
1;  
DB 2 TV 3

RESULT 11  
ID P83073 PRELIMINARY; PRT; 5 AA.  
AC P83073;  
DT 01-OCT-2001 (TREMBlrel. 18, Created)  
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
DE 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
DE 88 KDA PROTEIN (FRAGMENT).  
OS Bacillus cereus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1396;  
RN [1]  
RP SEQUENCE.  
RC STRAIN=NCIMB 11796;  
RA Browne N., Dowds B.C.A.;  
RL Submitted (JUL-2001) to the SWISS-PROT data bank.  
FT NON\_TER  
SQ SEQUENCE 5 AA; 623 MW; 6B01AAA336F00000 CRC64;

Query Match 20.0%; Score 7; DB 2; Length 5;  
Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TD 2  
1;  
DB 4 TE 5

RESULT 12  
ID 047029 PRELIMINARY; PRT; 7 AA.  
AC 047029;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE AAD A1 PROTEIN (FRAGMENT).

GN AAD A1.  
OS Enterobacter cloacae.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Enterobacter.  
OX NCBI\_TaxID=550;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94079349; PubMed=8257126;  
RA Kather P.N., Mann P.A., Mierwa R., Hare R.S., Miller G.H., Shaw K.J.;  
RT "Analysis of the aac(3)-Ia gene encoding a novel 3'-N-  
RT acetyltransferase.";  
RL Antimicrob. Agents Chemother. 37:2074-2079(1993).  
DR EMBL: M88012; AAA16193.1; -.  
FT NON\_TER  
SQ SEQUENCE 7 AA; 744 MW; 633862D2C321A030 CRC64;

Query Match 20.0%; Score 7; DB 2; Length 7;  
Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 5 LT 6  
1;  
DB 1 TT 2

RESULT 13  
ID P72081 PRELIMINARY; PRT; 7 AA.  
AC P72081;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE 3'-METHYLCEPHEM HYDROXYLASE (FRAGMENT).  
GN CEPE.  
OS Nocardia lactamdurans.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Pseudonocardiales; Pseudonocardaceae; Amycolatopsids.  
OX NCBI\_TaxID=1913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96009872; PubMed=7597411;  
RA Coque J., Perez-Llarena F.J., Enguita F.J., Fuente J.L., Martin J.F.,  
RA Liras P.;  
RT "Characterization of the cmcH genes of Nocardia lactamdurans and  
RT Streptomyces clavuligerus encoding a functional 3'-hydroxymethylcephem  
RT O-glycanoyltransferase for cephamycin biosynthesis.";  
RL Gene 162:21-27(1995).  
DR EMBL: Z21682; CAA79797.1; -.  
FT NON\_TER  
SQ SEQUENCE 7 AA; 746 MW; 75A1B2CDD1E681C0 CRC64;

Query Match 20.0%; Score 7; DB 2; Length 7;  
Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 TL 5  
1;  
DB 1 TM 2

RESULT 14  
ID 098866 PRELIMINARY; PRT; 7 AA.  
AC 098866;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
DE CYTOCHROME B/P SUBUNIT IV (FRAGMENT).  
OS Spinacia oleracea (Spinach).  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Splanactia.  
 CX NCBI\_TaxID=3562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-86120353; PubMed=3003688;  
 RA Sijben-Mueller G., Hallick R.B., Alt J., Westhoff P., Herrmann R.G.;  
 RT "Spinach plastid genes coding for initiation factor IF-1, ribosomal  
 RT protein S11 and RNA polymerase alpha-subunit.";  
 RL Nucleic Acids Res. 14:1029-1044(1986).  
 DR EMBL; X03496; CAA27215.1; -.  
 KW Chloroplast.  
 FT MON\_TER 1  
 SO SEQUENCE 7 AA; 907 MW; 644729D77409C420 CRC64;

Query Match 20.0%; Score 7; DB 8; Length 7;  
 Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
 Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 DE 3  
 : 1  
 Db 1 NF 2

RESULT 15  
 P82099 PRELIMINARY; PRT; 5 AA.  
 AC P82099;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE ELECTRIN 3.  
 OS Litoria rubella (Desert tree frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
 OC Litoria.  
 CX NCBI\_TaxID=104895;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE-SKIN SECRETION;  
 RA Wapnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;  
 RT "Peptides from the skin glands of the Australian buzzing tree frog  
 RT Litoria electrica. Comparison with the skin peptides from Litoria  
 RT rubella.";  
 RL Aust. J. Chem. 52:0-0(1999).  
 KW Amphibian skin; Amidation.  
 FT MOD\_RES 5  
 SO SEQUENCE 5 AA; 630 MW; 668761F2C9A0D000 CRC64;

Query Match 17.1%; Score 6; DB 13; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 F 3  
 : 1  
 Db 1 F 1

Search completed: July 15, 2002, 13:37:39  
 Job time: 729 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:25:26 ; Search time 107.71 Seconds

(without alignments)

7.219 Million cell updates/sec

Title: US-09-712-819A-5

Perfect score: 35

Sequence: 1 TDFITVIT 7

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 52936

Minimum DB seq length: 0

Maximum DB seq length: 7

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_032802:\*

1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:\*

2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:\*

3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:\*

4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:\*

5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:\*

6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:\*

7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:\*

8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:\*

9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:\*

10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:\*

11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:\*

12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:\*

13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:\*

14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:\*

15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:\*

16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:\*

17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:\*

18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:\*

19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:\*

20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:\*

21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:\*

22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	85.7	7	20	AAV40738
2	30	85.7	7	21	AAV40738
3	27	77.1	7	20	AAV40737
4	27	77.1	7	21	AAV40737
5	26	74.3	7	20	AAV40736
6	26	74.3	7	21	AAV40736
7	21	60.0	7	20	AAV40735
8	20	57.1	6	19	AAV40734
9	19	54.3	5	9	AAV40733
10	19	54.3	6	20	AAV40732
11	19	54.3	6	21	AAV40731

12	19	54.3	6	22	AAV40730
13	19	54.3	6	22	AAV40729
14	19	54.3	6	22	AAV40728
15	19	54.3	6	22	AAV40727
16	19	54.3	6	22	AAV40726
17	18	51.4	6	19	AAV40725
18	18	51.4	6	19	AAV40724
19	18	51.4	6	19	AAV40723
20	18	51.4	6	19	AAV40722
21	18	51.4	6	19	AAV40721
22	18	51.4	6	19	AAV40720
23	18	51.4	6	19	AAV40719
24	18	51.4	6	19	AAV40718
25	18	51.4	6	19	AAV40717
26	18	51.4	6	19	AAV40716
27	18	51.4	6	19	AAV40715
28	18	51.4	6	19	AAV40714
29	18	51.4	6	19	AAV40713
30	18	51.4	6	19	AAV40712
31	18	51.4	6	19	AAV40711
32	18	51.4	6	19	AAV40710
33	18	51.4	6	19	AAV40709
34	18	51.4	6	19	AAV40708
35	18	51.4	6	19	AAV40707
36	18	51.4	6	19	AAV40706
37	18	51.4	6	19	AAV40705
38	18	51.4	6	19	AAV40704
39	18	51.4	6	19	AAV40703
40	18	51.4	6	19	AAV40702
41	18	51.4	6	19	AAV40701
42	18	51.4	6	19	AAV40700
43	18	51.4	6	19	AAV40699
44	18	51.4	6	19	AAV40698
45	18	51.4	6	19	AAV40697

#### ALIGNMENTS

RESULT 1

AAV40738

ID AAV40738 standard; peptide: 7 AA.

XX

AC AAV40738:

XX

DT 01-DEC-1999 (first entry)

XX

DE S4 derivative #12, beta strand of scaffold protein structure.

XX

KW Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;

XX

OS Synthetic.

XX

PN EP947582-A1.

XX

PD 06-OCT-1999.

XX

PF 31-MAR-1998; 98EP-0870065.

XX

PR 31-MAR-1998; 98EP-0870065.

XX

PA (INNO-) INNOGENETICS NV.

XX

PI Desmet J, Bufton S, Hoogenboom H, Sablon E;

XX

DR WPI; 1999-542958/46.

XX

PT New scaffold protein, useful for stabilizing antigens used as vaccines

XX

PS Disclosure: Page 6; 105pp; English.

XX

Epitope tag #4 use

Integrin activator

AUS epitope used 1

AUS peptide epitop

Amaraanthus vitidis

Hexapeptide #13 bl

Hexapeptide #13 bl

LFA-1 alpha subunit

Trypatic 40 KD subu

Peptide CDR-H1-7 d

Antihaptilis C pe

Human anti-HBs ant

IL-1 antagonist pe

AAV VP3 derived pe

Hepatooma diagnosi

Epidermal growth f

Human pancreatic p

Human SNP associat

Hepatooma-diagnosi

T cell surface rec

CD90 C-terminal co

Ribonuclease reduc

Human beta1.6-N-ac

T-cell epitope #3

Antimalarial pepti

Antimalarial pepti

Antimalarial pepti

Antimalarial pepti

S3 derivative #20

S4 derivative #9,

Liquid interfeon-

Scaffold protein S

Scaffold protein S

AAV02727-7400748 are functionally equivalent derivatives of the S4 peptide (AAV046067) which forms part of a scaffold protein. S4 is a beta strand peptide which forms part of a beta sheet. Peptides (AAV046061-740609) together form a single-chain scaffold protein which contains at least 1 disulfide bond, contains less than 10% alpha helix and contains at least 6 beta-strands. The scaffold protein is constructed of beta strands S1-56, and may also include beta strands A1-A3, or any functionally equivalent derivative of these sequences. The beta strands form two beta sheets S1/S4/S3 and S6/S5/S2 with each strand connected to the next by hydrogen bonds, which generate a beta sandwich architecture. If the additional beta strands A1-43 are included in the structure the scaffold is constructed of two beta sheets, with the structures A1/S1/S4/S3 and S6/S5/S2/S2/3. The beta strands are connected to each other via amino acid loops, where at least one of the loops binds to a receptor or antigen. The scaffold protein is used to stabilize antigens or whole proteins such as receptors, or their fragments. It may be used to bind two separate molecules. For example, one surface of the scaffold may be bound to a protein which binds to a tumour antigen. This will target the complex to tumour cells. Another surface may be bound to a cytotoxic molecule or an autoimmune antibody which may then kill the tumour cells. Therefore the scaffold protein may be used to target chemotherapeutic agents to specific cells. It may also be used to stabilize individual peptides in a peptide library and may be used in diagnostic techniques, and to stabilize antigens used as vaccines.

Query Match	85.7%;	Score 30;	DB 20;	Length 7;
Best Local Similarity	100.0%;	Pred. No. 6.4e+05;		
Matches	6;	Conservative	0;	Indels 0;
		Mismatches	0;	Gaps 0;

QY	2	DEFLTY	7
Db	1	deflty	6

RESULT	2
AAB30076	
ID	AAB30076 standard; Peptide; 7 AA.

AC	AAB30076;
XX	
DT	09-FEB-2001 (first entry)

DE Scaffold protein SCA 54 peptide SEQ ID NO: 137.

KM Human; CTLA-4; scaffold protein; antigen-binding; receptor-binding;  
 KM SCA domain; cancer; thrombosis; osteoporosis, rheumatoid arthritis;  
 KM diabetic retinopathy; atherosclerosis.

OS Synthetic.

PN WO2000060070-A1.

PD 12-OCT-2000.

PF 01-APR-1999; 99WO-EP02283.

01-APR-1999; 99WO-EP02283.

PA (INNO-) INNOGENETICS NV,

Desmet J, Hufton S, Hoogenboom H, Sablon E;

DR WPT; 2000-665002/64.

PT scaffold composed of single-chain polypeptide having beta sandwich  
PT architecture carrying new and randomized peptide sequences useful as  
PT supporting framework and carrying antigen- or receptor binding

PS Disclosure; page 15; 68pp; English.

xx The present invention is concerned with producing scaffold proteins  
xx based upon the human CTLA-4 SCM domain. These scaffold proteins can be  
xx used as a scaffold to bind antigen- or receptor-binding fragments. These  
xx can be used in the treatment of diseases such as cancer,  
xx atherosclerosis, thrombosis, osteoporosis, rheumatoid arthritis and  
xx diabetic retinopathy. Sequences AAB2930-B2939 were used in the  
xx production of the proteins of the invention.

xx Sequence 7 AA:

Sequence 7 AA:

Query Match	85.78; Score 30; DB 21; Length 7;
-------------	-----------------------------------

Oy	2	DEFTI	7
	1	11111	
Db	1	dftlci	6

RESULT	3
AAY40737	
ID	AAY40737 standard; peptide; 7 AA.

AC AA40737;

DT 01-DEC-1999 (first entry)

DE S4 derivative #11, beta strand of scaffold protein structure.

\*\*\* Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;  
KW tumour; chemotherapeutic agent.

OS Synthetic.

PN EP947582-A1.

PD 06-OCT-1999.

PF 31-MAR-1998;

PR 31-MAR-1998;

PA (INNO-) INNOGENETICS NV.

Desmet J, Hufton S, Hoogenboom H, Sablon E; PI

DR WPI; 1999-542958/46.

PT New scaffold protein, useful for stabilizing antigens used as vaccines

EXX

CC Sequences (AAV40727-5407748 are functionally equivalent derivatives of the  
CC S4 peptide (AAV406007) which forms part of a scaffold protein. S4 is a  
CC beta strand peptide which forms part of a beta sheet. Peptides  
CC beta15601(Y40603) together form a single-chain scaffold protein which  
CC contains at least 1 disulfide bond, contains less than 10% alpha helix  
CC and contains at least 6 beta-strands. The scaffold protein is constructed  
CC of beta strands S1-S6, and may also include beta strands A1-A3, or any  
CC functionally equivalent derivative of these sequences. The beta strands  
CC form two beta sheets S1/S4/S3 and S6/S5/S2 with each strand connected to  
CC the next by hydrogen bonds, which generate a beta sandwich architecture.  
CC If the additional beta strands A1-A3 are included in the structure the  
CC scaffold is constructed of two beta sheets, with the structures  
CC A1/S1/S4/S3 and S6/S5/S2/A2/A3. The beta strands are connected to each  
CC other via amino acid loops, where at least one of the loops binds to a  
CC receptor or antigen. The scaffold protein is used to stabilize antigens  
CC or whole proteins such as receptors, or their fragments. It may be used  
CC to bind two separate molecules. For example, one surface of the scaffold  
CC may be bound to a protein which binds to a tumour antigen. This will

CC target the complex to tumour cells. Another surface may be bound to a  
CC cytotoxic molecule or an autoimmune antibody which may then kill the  
CC tumour cells. Therefore the scaffold protein may be used to target  
CC chemotherapeutic agents to specific cells. It may also be used to  
CC stabilize individual peptides in a peptide library and may be used in  
CC diagnostic techniques, and to stabilize antigens used as vaccines.

XX  
SQ Sequence 7 AA:

Query Match 77.1%; Score 27; DB 20; Length 7;  
Best Local Similarity 83.3%; Pred. No. 6.4e+05;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 DPTLTI 7  
1:|||||  
Db 1 dyltli 6

RESULT 4

AAB30075 AAB30075 standard; Peptide; 7 AA.

XX  
AC AAB30075;

DT 09-FEB-2001 (first entry)

XX Scaffold protein SCA S4 peptide SEQ ID NO: 136.

DE Human; CTLA-4; scaffold protein; antigen-binding; receptor-binding;  
KW SCA domain; cancer; thrombosis; osteoporosis, rheumatoid arthritis;  
KV diabetic retinopathy; atherosclerosis.

XX Synthetic.

OS  
XX WO200060070-A1.

PN 12-OCT-2000.

XX  
PD 01-APR-1999; 99MO-EP02283.

XX  
PE 01-APR-1999; 99MO-EP02283.

XX (INNO-) INNOGENETICS NV.

PA Desmet J, Hufon S, Hoogenboom H, Sablon E;

PI WPI: 2000-665002/64.

DR Scaffolds composed of single-chain polypeptide having beta sandwich  
XX architecture carrying new and randomized peptide sequences useful as

PT supporting framework and carrying antigen- or receptor binding

PT fragments -

XX Disclosure: Page 15; 68pp; English.

XX The present invention is concerned with producing scaffold proteins  
CC based upon the human CTLA-4 SCA domain. These scaffold proteins can be  
CC used as a scaffold to bind antigen- or receptor-binding fragments. These  
CC can be used in the treatment of diseases such as cancer,  
CC atherosclerosis, thrombosis, osteoporosis, rheumatoid arthritis and  
CC diabetic retinopathy. Sequences AAB29930-B29939 were used in the  
CC production of the proteins of the invention.

XX  
SQ Sequence 7 AA:

Query Match 77.1%; Score 27; DB 21; Length 7;

Best Local Similarity 83.3%; Pred. No. 6.4e+05;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 DPTLTI 7  
1:|||||

Db 1 dyltli 6

RESULT 5

AAY40736 AAY40736 standard; peptide; 7 AA.

XX  
AC AAY40736;

DT 01-DEC-1999 (first entry)

XX S4 derivative #10, beta strand of scaffold protein structure.

DE Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;

KW tumour; chemotherapeutic agent.

XX Synthetic.

OS  
XX EP947582-A1.

PN 06-OCT-1999.

XX  
PD 31-MAR-1998; 98EP-0870065.

XX  
PE 31-MAR-1998; 98EP-0870065.

XX (INNO-) INNOGENETICS NV.

PA Desmet J, Hufon S, Hoogenboom H, Sablon E;

PI WPI: 1999-542958/46.

DR New scaffold protein, useful for stabilizing antigens used as vaccines

XX  
PT  
XX Disclosure: Page 6; 105pp; English.

PS Sequences AAY40727-YA0748 are functionally equivalent derivatives of the

CC S4 peptide (AAY40607) which forms part of a scaffold protein. S4 is a

CC beta strand peptide which forms part of a beta sheet. Peptides

CC (AAY40601-YA0609) together form a single-chain scaffold protein which

CC contains at least 1 disulfide bond, contains less than 10% alpha helix

CC and contains at least 6 beta-strands. The scaffold protein is constructed

CC of beta strands S1-S6, and may also include beta strands A1-A3, or any

CC functionally equivalent derivative of these sequences. The beta strands

CC form two beta sheets S1/S4/S3 and S6/S5/S2 with each strand connected to

CC the next by hydrogen bonds, which generate a beta sandwich architecture.

CC If the additional beta strands A1-A3 are included in the structure the

CC scaffold is constructed of two beta sheets, with the structures

CC A1/S1/S4/S3 and S6/S5/S2/A2/A3. The beta strands are connected to each

CC other via amino acid loops, where at least one of the loops binds to a

CC receptor or antigen. The scaffold protein is used to stabilize antigens

CC or whole proteins such as receptors, or their fragments. It may be used

CC to bind two separate molecules. For example, one surface of the scaffold

CC target the complex to tumour cells. Another surface may be bound to a

CC cytotoxic molecule or an autoimmune antibody which may then kill the

CC tumour cells. Therefore the scaffold protein may be used to target

CC chemotherapeutic agents to specific cells. It may also be used to

CC stabilize individual peptides in a peptide library and may be used in

CC diagnostic techniques, and to stabilize antigens used as vaccines.

XX  
SQ Sequence 7 AA:

Query Match 74.3%; Score 26; DB 20; Length 7;  
Best Local Similarity 83.3%; Pred. No. 6.4e+05;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 DPTLTI 7  
1:|||||  
Db 1 dyltli 6

RESULT 6  
AAB30074  
ID AAB30074 standard; Peptide: 7 AA.  
XX  
AC AAB30074;  
XX  
DT 09-FEB-2001 (first entry)  
XX  
DE Scaffold protein SCA s4 peptide SEQ ID NO: 135.  
XX  
KW Human; CTR4-4; scaffold protein; antigen-binding; receptor-binding;  
KW SCA domain; cancer; thrombosis; osteoporosis; rheumatoid arthritis;  
KW diabetic retinopathy; atherosclerosis.  
XX  
OS Synthetic.  
XX  
PN WO20006070-A1.  
XX  
PD 12-OCT-2000.  
XX  
PE 01-APR-1999; 99WO-EP02283.  
XX  
PR 01-APR-1999; 99WO-EP02283.  
XX  
PA (INNO-) INNOGENETICS NV.  
XX  
PI Desmet J, Hufion S, Hoogenboom H, Sablon E;  
XX  
DR WPI: 2000-665002/64.  
XX  
PT Scaffold composed of single-chain polypeptide having beta sandwich  
PT architecture carrying new and randomized peptide sequences useful as  
PT supporting framework and carrying antigen- or receptor binding  
XX  
XX  
XX  
XX  
PS Disclosure: Page 15; 68pp; English.  
XX  
CC The present invention is concerned with producing scaffold proteins  
CC based upon the human CTR4-4 SCA domain. These scaffold proteins can be  
CC used as a scaffold to bind antigen- or receptor-binding fragments. These  
CC can be used in the treatment of diseases such as cancer,  
CC atherosclerosis, thrombosis, osteoporosis, rheumatoid arthritis and  
CC diabetic retinopathy. Sequences AAB29930-829939 were used in the  
CC production of the proteins of the invention.  
XX  
SQ Sequence 7 AA:  
  
Query Match 74.3%; Score 26; DB 21; Length 7;  
Best Local Similarity 83.3%; Pred. NO. 6.4e+05;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 DFTLTI 7  
DB 1 dftlsl 6  
  
RESULT 7  
AAV14399  
ID AAV14399 standard; peptide: 7 AA.  
XX  
AC AAV14399;  
XX  
DT 17-AUG-1999 (first entry)  
XX  
DE Peptide CDR-H1-7(Y3F) derived from anti-HCV protease MAb 8D4.  
XX  
KW Complementarity determining region; CDR; monoclonal antibody; MAb;  
KW hepatitis C virus; HCV; protease; binding site.  
XX  
OS Synthetic.

PN JP11127861-A.  
XX  
PD 18-MAY-1999.  
XX  
PE 29-OCT-1997; 97JP-0297451.  
XX  
PR 29-OCT-1997; 97JP-0297451.  
XX  
PA (NIIHA) JAPAN ENERGY CORP.  
XX  
DR WPI: 1999-350322/30.  
XX  
PT Neutralized antibody partial peptide derived from hepatitis C virus  
PT - useful for inhibiting hepatitis C virus (HCV) serine protease  
PT activity  
XX  
PS Example 1; Page 24; 32pp; Japanese.  
XX  
CC This sequence corresponds to a peptide (CDR-H1-7; AAV14403) derived from  
CC the sequence of the heavy chain variable region complementarity  
CC determining region (CDR)-1 of the anti-hepatitis C virus (HCV) ser/Thr  
CC protease monoclonal antibody (MAb) 8D4 protein. The peptide has a Tyr  
CC to Phe amino acid substitution at position 3 compared to the CDR-H1-7  
CC peptide. The invention relates to the use of partial peptides  
CC (AAV14348-Y14353) from the MAb 8D4 for inhibiting HCV serine protease  
CC activity.  
XX  
SQ Sequence 7 AA:  
  
Query Match 60.0%; Score 21; DB 20; Length 7;  
Best Local Similarity 80.0%; Pred. NO. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 TDFTL 5  
DB 1 tdfvl 5  
  
RESULT 8  
AAW31467  
ID AAW31467 standard; Protein: 6 AA.  
XX  
AC AAW31467;  
XX  
DT 04-AUG-1998 (first entry)  
XX  
DE Transcriptional activator peptide fragment 1S130.  
XX  
KW Activating sequence; Gal4; transcriptional activator; RNA polymerase;  
KW Protein-protein interaction; gene therapy; therapeutic; holoenzyme;  
KW Gal11; DNA binding domain.  
XX  
OS Synthetic.  
XX  
PN WO9744447-A2.  
XX  
PD 27-NOV-1997.  
XX  
PE 02-MAY-1997; 97WO-US07338.  
XX  
PR 01-MAY-1997; 97US-0017016.  
PR 03-MAY-1996; 96US-0017016.  
XX  
PA (HARD) HARVARD COLLEGE.  
XX  
PL Lu X, Peashine M, Wu Y;  
XX  
DR WPI: 1998-018502/02.  
DR N-PSDB; AAV02565.  
XX  
PT New transcriptional activator containing DNA binding domain bound to  
PT peptide - useful for controlling gene expression, especially in gene



PT therapy, and in protein-protein interaction assays, does not inhibit  
PT other transcription activators  
PS Example 1; Page 26; 55pp; English.  
XX  
CC AAM31412-W31459, AAM31461-W31481, AAM31483-W31493 and AAM39053-W39076  
CC are fragments used in an assay to determine novel transcriptional  
CC activators. The method involves the production of transcriptional  
CC activators comprising of a DNA-binding group and a 6-25 amino acid  
CC peptide that is covalently bonded to the DNA binding group and does not  
CC represent a fragment of a natural transcription activator.  
CC Protein-protein interactions are identified in the assay by fusing a  
CC DNA-binding domain to a library of DNA fragments and introducing this and  
CC a fusion of target protein and a polypeptide containing a region of Gal4  
CC which interacts with Gal1p into a cell containing Gal1p and identifying  
CC members of the library that interact with the target transcription in a  
CC cell, e.g. for controlling gene activity, particularly in gene therapy  
CC (e.g. recognizing a site close to a selected therapeutic gene).  
CC Transcription can be activated without blocking other transcriptional  
CC activators. They probably act by interacting with a component of the RNA  
CC polymerase II holoenzyme, Gal11, the strongest known yeast activator,  
CC which provides a more sensitive assay allowing detection of even weak  
CC protein-protein interactions. Such activators do not create toxicity  
CC problems even when overexpressed.  
XX  
SO Sequence 6 AA:  
  
Query Match 57.1%; Score 20; DB 19; Length 6;  
Best Local Similarity 80.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
OY 1 TDFTL 5  
|||  
|  
Db 1 tdfll 5  
  
RESULT 9  
AAP82200  
ID AAP82200 standard; protein; 5 AA.  
XX  
AC AAP82200;  
XX  
DT 25-OCT-1990 (first entry)  
XX  
DE Example of peptide 2 for treatment of schizophrenia or psoriasis.  
XX  
KM schizophrenia; psoriasis; vasoactive intestinal polypeptide.  
XX  
OS synthetic.  
XX  
FN SE8700125-A.  
XX  
PD 16-JUL-1988.  
XX  
PF 15-JAN-1987; 87SE-0000125.  
XX  
PR 15-JAN-1987; 87SE-0000125.  
XX  
PA (WETT/) WETTERBERG.  
XX  
PI Wetterberg L;  
XX  
DR WPI; 1988-328337/46.  
XX  
XX Short peptides for treatment of psoriasis and schizophrenia -  
PT comprise vasoactive intestinal polypeptide, peptide T or short  
PT peptide with five aminoacids  
XX  
PS Claim 1; Page 5; 9pp; Swedish.  
XX  
CC Specific example of pentapeptide of the general formula of AAP82197.

CC These peptides can be administered intravenously,  
CC topically or perorally to relieve the symptoms of psoriasis or  
CC schizophrenia. Amino acids at posns 2 and 3 can be any residue but  
CC Asp is preferred at posn 3. See also AAP82196-9.  
XX  
SO Sequence 5 AA:  
  
Query Match 54.3%; Score 19; DB 9; Length 5;  
Best Local Similarity 75.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 TDFTL 4  
|||  
|  
Db 2 tdyt 5  
  
RESULT 10  
AAU06466  
ID AAU06466 standard; Peptide; 6 AA.  
XX  
AC AAU06466;  
XX  
DT 27-SEP-1999 (first entry)  
XX  
DE Epitope tag.  
XX  
KW Epitope tag; antibody engineering; yeast; surface display;  
KW protein library; peptide library.  
XX  
OS Synthetic.  
XX  
PN WO9936569-A1.  
XX  
PD 22-JUL-1999.  
XX  
PF 20-JAN-1999; 99WO-US01188.  
XX  
PR 26-AUG-1998; 98US-0140084.  
XX  
PR 20-JAN-1998; 98US-0009388.  
XX  
PA (UNII ) UNIV ILLINOIS FOUND.  
XX  
PI Boder ET, Kieke MC, Kranz DM, Shusta E, Wiltrop KD;  
XX  
DR WPI; 1999-430619/36.  
XX  
XX Selecting proteins with enhanced phenotypic properties than  
PT wild-type proteins, is useful for highly specific cancer diagnosis  
PT and therapy  
PS Disclosure; Page 7; 116pp; English.  
XX  
XX This peptide comprises an epitope tag that can be used in  
CC methods of the invention. The invention discloses a powerful new  
CC system for engineering antibody affinity and specificity, by  
CC constructing a microbial analogue of the mammalian system's B cell  
CC repertoire. Antibodies are displayed on the surface of yeast cells  
CC by genetic fusion with yeast cell wall proteins, especially  
CC agglutinin proteins. After mutation, variants are selected on the  
CC basis of improved binding characteristics with fluorescently  
CC labeled targets. The selection method also identifies proteins  
CC with enhanced phenotypic characteristics, proteins that are  
CC displayed at higher levels, proteins that are secreted at higher  
CC efficiency and proteins of improved stability.  
XX  
SO Sequence 6 AA:  
  
Query Match 54.3%; Score 19; DB 20; Length 6;  
Best Local Similarity 80.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TDFYL 5  
DB 1 tdfyl 5

## RESULT 11

AAE13076 standard; peptide; 6 AA.  
ID AAY77708

AC AAY77708;

DT 12-MAY-2000 (first entry)

DE A05 peptide epitope.

KW Cell surface receptor; luminescence; protein internalization;  
drug discovery; screening assay; epitope; A05.

OS Synthetic.

PN W0200003246-A2.

PD 20-JAN-2000.

PF 13-JUL-1999; 99WO-US15870.

PR 13-JUL-1998; 98US-0092671.

PA (CELL-) CELLOMICS INC.

PI Rubin RA, Giuliano KA, Gough A, Dunlay T;

DR WPI; 2000-171170/15.

PT Automated screening method for identifying compounds which induce cell  
surface receptor internalization, useful for drug discovery -

PS Example 6; Page 67; 148pp; English.

CC The invention relates to a method for identifying compounds which  
inhibit internalization of cell surface receptors. Provided are an array  
of locations, each containing cells with a cell surface receptor  
protein, that are treated with a test compound. The protein is  
luminescently labeled or contacted with a luminescently labeled cell  
before or after test compound treatment. Any luminescence produced is  
converted into digital data and automatically analysed to determine if  
the test compound induced the protein internalization. The novel method  
is used to screen for compounds which modulate cell surface receptor  
protein internalization, this can be used in drug discovery, to test  
compound efficacy in living biological systems. The assay method is  
automated and compact. It has high throughput and uses smaller volumes of  
reagents and test compounds. Sequences AAY77704-718 represent examples of  
peptide epitope tags used in the course of the invention.

SQ Sequence 6 AA;

Query Match 54.3%; Score 19; DB 21; Length 6;  
Best Local Similarity 80.0%; Pred. No. 6.4e+05;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TDFYL 5  
DB 1 tdfyl 5

## RESULT 12

AAE13076 standard; peptide; 6 AA.  
ID AAE13076

AC AAE13076;

DT 28-JAN-2002 (first entry)

XX Epitope tag #4 used in yeast cell surface display of proteins.

DE Phenotypic property; yeast; cell wall protein; epitope tag.

OS Unidentified.

PN US6300065-B1.

PD 09-OCT-2001.

PF 26-AUG-1998; 98US-0140084.

PR 31-MAY-1996; 96US-018741P.

PR 30-MAY-1997; 97US-0866398.

PR 20-JAN-1998; 98US-0009388.

PA (UNIT ) UNIV ILLINOIS FOUND.

PI Kieke MC, Wiltrop KD, Boder ET, Kranz DW, Shusta E;

DR WPI; 2001-656236/75.

PT Selecting proteins, e.g. antibodies, with enhanced phenotypic  
properties relative to those of a wild-type comprises transforming  
yeast cells with a vector expressing a protein to be tested fused to a  
yeast cell wall protein -

PS Disclosure; Column 5; 64pp; English.

CC The present invention relates to a method for selecting proteins with  
enhanced phenotypic properties relative to those of a wild-type.  
CC comprises transforming yeast cells with a vector expressing a protein  
to be tested fused to a yeast cell wall protein. The method is  
particularly useful for selecting antibodies for improved affinity and  
specificity. The present sequence is an epitope tag which is used in  
yeast cell surface display of proteins.

SQ Sequence 6 AA;

Query Match 54.3%; Score 19; DB 22; Length 6;  
Best Local Similarity 80.0%; Pred. No. 6.4e+05;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TDFYL 5  
DB 1 tdfyl 5

## RESULT 13

AAE13076 standard; peptide; 6 AA.  
ID AAE13076

AC AAE13076;

DT 08-JAN-2002 (first entry)

DE Integrin activating peptide SEQ ID NO 1.

KW Integrin stimulant; vulnarary; injury healing;  
postsurgical tissue recovery.

OS Unidentified.

PN JP2001213898-A.

PD 07-AUG-2001.

PR 31-JAN-2000; 2000JP-0022469.

PR 31-JAN-2000; 2000JP-0022469.



Mon Jul 15 13:58:59 2002

us-09-712-819a-5.closed.rag

Page 8

Db 1 tdfy 5

Search completed: July 15, 2002, 13:25:27  
Job time: 1458 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:26:14 ; Search time 40.89 seconds  
(without alignments)  
4.181 Million cell updates/sec

Title: US-09-712-819A-5  
Perfect score: 35  
Sequence: 1 TDFLTI 7  
Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 39160

Minimum DB seq length: 0  
Maximum DB seq length: 7

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCRTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/Backfilest.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	54.3	6	4	US-09-140-084-4
2	19	54.3	7	3	US-08-916-443A-8
3	19	54.3	7	4	US-08-640-737-38
4	17	48.6	5	1	US-08-405-230-10
5	17	48.6	7	1	US-08-910-990-10
6	17	48.6	7	1	US-08-136-743B-29
7	17	48.6	7	1	US-08-136-743B-30
8	17	48.6	7	1	US-08-136-743B-31
9	17	48.6	7	1	US-08-136-743B-32
10	17	48.6	7	1	US-08-405-230-5
11	17	48.6	7	2	US-08-910-990-5
12	17	48.6	7	2	US-08-739-401A-6
13	17	48.6	7	5	PCT-US93-11703-72
14	16	45.7	5	1	US-08-180-209B-14
15	16	45.7	5	1	US-08-385-745-14
16	16	45.7	5	4	US-08-385-745-14
17	16	45.7	5	4	US-08-591-632-17
18	16	45.7	5	4	US-08-591-632-23
19	16	45.7	5	4	US-08-591-632-26
20	16	45.7	5	4	US-08-474-853-14
21	16	45.7	5	5	PCT-US94-02629-14
22	16	45.7	6	1	US-08-252-995D-7
23	16	45.7	6	2	US-08-482-228-180
24	16	45.7	6	2	US-08-834-108-7
25	16	45.7	6	3	US-08-482-528-180
26	16	45.7	7	1	US-08-136-743B-55
27	16	45.7	7	1	US-08-096-946-5

28	16	45.7	7	2	US-08-177-109A-7	Sequence 7, Appl
29	16	45.7	7	2	US-08-687-706-7	Sequence 7, Appl
30	16	45.7	7	3	US-09-040-216-28	Sequence 28, Appl
31	16	45.7	7	4	US-09-173-641-52	Sequence 52, Appl
32	16	45.7	7	5	PCT-US94-07329-5	Sequence 5, Appl
33	15	42.9	5	1	US-08-136-743B-63	Sequence 63, Appl
34	15	42.9	5	1	US-07-789-184-126	Sequence 126, App
35	15	42.9	5	1	US-08-475-263-126	Sequence 126, App
36	15	42.9	5	1	US-08-485-886-126	Sequence 126, App
37	15	42.9	5	2	US-08-667-001-122	Sequence 126, App
38	15	42.9	5	2	US-08-477-362-126	Sequence 126, App
39	15	42.9	5	2	US-08-477-134-126	Sequence 126, App
40	15	42.9	5	3	US-08-473-489A-126	Sequence 126, App
41	15	42.9	5	3	US-09-040-216-55	Sequence 55, Appl
42	15	42.9	5	3	US-08-485-695-126	Sequence 126, App
43	15	42.9	5	4	US-08-018-760-126	Sequence 126, App
44	15	42.9	6	1	US-08-136-743B-62	Sequence 62, Appl
45	15	42.9	6	1	US-08-290-448A-41	Sequence 41, Appl

## ALIGNMENTS

RESULT 1  
US-09-140-084-4  
Sequence 4, Application US/09140084A  
Patent No. 6300065  
GENERAL INFORMATION:  
APPLICANT: Kiege, et al.  
TITLE OF INVENTION: Yeast Cell Surface Display of Proteins and Uses Thereof  
FILE REFERENCE: D6061CIP2  
CURRENT FILING DATE: 1998-08-26  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 6  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: Epitope Tag  
US-09-140-084-4

Query Match 54.3%; Score 19; DB 4; Length 6;  
Best Local Similarity 80.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TDFLTI 5  
DB 1 TDFLTI 5

RESULT 2  
US-08-916-443A-8  
Sequence 8, Application US/08916443A  
Patent No. 6001986  
GENERAL INFORMATION:  
APPLICANT: Yong Sig KIM  
APPLICANT: Sun Chung PARK  
APPLICANT: Soo Kyung OH  
APPLICANT: Hosuli LEE  
APPLICANT: Jeong Woo CHO  
APPLICANT: Chang H. CHUNG  
TITLE OF INVENTION: Antiviral Proteins, Amaranth 1 and 2, from  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenyon & Kenyon  
STREET: 1025 Connecticut Avenue, N.W., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA

ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3+ floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
SOFTWARE: WordPerfect 6.1 Windows  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/916,443A  
FILING DATE: 22 AUG 1997  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Toffenetti, Judith L.  
REGISTRATION NUMBER: 39,048  
REFERENCE/DOCKET NUMBER: 1942/18  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-429-1776  
TELEFAX: 202-429-0796  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-916-443A-8

Query Match 54.3%; Score 19; DB 3; Length 7;  
Best Local Similarity 50.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 DFTLTI 7  
1 1 1  
DB 2 DFTFTV 7

RESULT 3  
US-08-640-737-38

Sequence 38, Application US/08640737  
Patent No. 6215044  
GENERAL INFORMATION:  
APPLICANT: ARROWSMITH, David A.  
APPLICANT: HELLER, Susan A.  
APPLICANT: DE SILVA, Jacqueline  
APPLICANT: WHITEMAN, Sally A.  
TITLE OF INVENTION: Tomato Xyloglucan Endo-Transglycosylase  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.  
STREET: 1100 New York Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MS Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/640,737  
FILING DATE: 06-MAY-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB94/02467  
FILING DATE: 10-NOV-1994  
PRIOR APPLICATION NUMBER: GB 9323225.4  
FILING DATE: 10-NOV-1993  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown

US-08-640-737-38

Query Match 54.3%; Score 19; DB 4; Length 7;  
Best Local Similarity 80.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 DFTLTI 6  
1 1 1  
DB 3 DFTFTV 7

RESULT 4  
US-08-405-230-10

Sequence 10, Application US/08405230  
Patent No. 5707846  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
APPLICANT: NISHIKAWA, Aetsushi  
APPLICANT: YAMAGUCHI, No. 5707846om1  
TITLE OF INVENTION: NOVEL N-ACETYLGUCOSAMINYL TRANSFERASE  
TITLE OF INVENTION: GENE CODING THEREFOR AND PROCESS FOR PRODUCTION THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/405,230  
FILING DATE: 16-MAR-1995  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,736

FILING DATE: 23-AUG-1993  
APPLICATION NUMBER: JP 4-245950  
FILING DATE: 24-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-237118

FILING DATE: 06-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 001560-215  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:

NAME/KEY: Peptide  
LOCATION: 1..5  
OTHER INFORMATION: /note="Amino acid sequence  
OTHER INFORMATION: encoded by nucleotides 1-15 of SEQ ID NO. 7."  
US-08-405-230-10

Query Match 48.6%; Score 17; DB 1; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDF 3  
111  
DB 2 TDF 4

RESULT 5  
US-08-910-990-10  
Sequence 10 Application US/08910990  
Patent No. 5834284  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
APPLICANT: NISHIKAWA, Atsushi  
APPLICANT: YAMAGUCHI, No. 5834284oml  
TITLE OF INVENTION: NOVEL N-ACETYLGLOUCOSAMINYL TRANSFERASE  
TITLE OF INVENTION: GENE CODING THEREFOR AND PROCESS FOR PRODUCTION THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22133-1404  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,990  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,230  
FILING DATE: 16-MAR-1995  
APPLICATION NUMBER: US 08/110,736  
FILING DATE: 23-AUG-1993  
APPLICATION NUMBER: JP 4-245950  
FILING DATE: 24-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-237118  
FILING DATE: 06-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 001560-215  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..5  
OTHER INFORMATION: /note="Amino acid sequence  
OTHER INFORMATION: encoded by nucleotides 1-15 of SEQ ID NO. 7."  
US-08-910-990-10  
Query Match 48.6%; Score 17; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TDF 3  
111  
DB 2 TDF 4

RESULT 6  
US-08-136-743B-29

Sequence 29 Application US/08136743B  
Patent No. 5459063

GENERAL INFORMATION:

APPLICANT: Barry S. Cooperman, Harvey Rubin,  
APPLICANT: Jerome Salem, and Allison L. Fisher

TITLE OF INVENTION: "Plasmodium falciparum Ribonu-

TITLE OF INVENTION: cleotide Reductase, DNA Sequences Therefor and Peptide inh1

TITLE OF INVENTION: Thereof"

NUMBER OF SEQUENCES: 67

CORRESPONDENCE ADDRESS:

ADDRESSEE: The University of Pennsylvania

STREET: Suite 330

STREET: 3700 Market Street

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: U.S.A.

ZIP: 19104-3246

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/136,743B

FILING DATE: 10/14/93

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Monaco, Daniel A.

REGISTRATION NUMBER: 30,480

REFERENCE/DOCKET NUMBER: 3957-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-8383

TELEFAX: (215) 568-5549

TELEX: No. 5459063e

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 7 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-136-743B-29

Query Match

48.6%; Score 17; DB 1; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;

Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDF 3  
111  
DB 5 TDF 7

RESULT 7  
US-08-136-743B-30

Sequence 30 Application US/08136743B  
Patent No. 5459063

GENERAL INFORMATION:

APPLICANT: Barry S. Cooperman, Harvey Rubin,  
APPLICANT: Jerome Salem, and Allison L. Fisher

TITLE OF INVENTION: "Plasmodium falciparum Ribonu-

TITLE OF INVENTION: cleotide Reductase, DNA Sequences Therefor and Peptide inh1

TITLE OF INVENTION: Thereof"

NUMBER OF SEQUENCES: 67

CORRESPONDENCE ADDRESS:

ADDRESSEE: The University of Pennsylvania

STREET: Suite 330

STREET: 3700 Market Street

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: U.S.A.

ZIP: 19104-3246  
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 Inch, 720 Kb  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/136,743B  
FILING DATE: 10/14/93  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Monaco, Daniel A.  
REGISTRATION NUMBER: 30,480  
REFERENCE/DOCKET NUMBER: 3957-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-8383  
TELEFAX: (215) 568-5549  
TELEX: NO. 5459063e  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-136-743B-30

Query Match 48.6%; Score 17; DB 1; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDF 3  
111  
DB 5 TDF 7

RESULT 8  
US-08-136-743B-31  
Sequence 31, Application US/08136743B  
Patent No. 5459063  
GENERAL INFORMATION:  
APPLICANT: Barry S. Cooperman, Harvey Rubin,  
Applicant: Jerome Salem, and Allison L. Fisher  
TITLE OF INVENTION: "Plasmodium falciparum Ribonu-  
TITLE OF INVENTION: cleotide Reductase, DNA Sequences Therefor and Peptide Inhibit  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The University of Pennsylvania  
STREET: Suite 330  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: U.S.A.  
ZIP: 19104-3246  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 Inch, 720 Kb  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/136,743B  
FILING DATE: 10/14/93  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Monaco, Daniel A.  
REGISTRATION NUMBER: 30,480  
REFERENCE/DOCKET NUMBER: 3957-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-8383  
TELEFAX: (215) 568-5549  
TELEX: NO. 5459063e  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
US-08-136-743B-31

Query Match 48.6%; Score 17; DB 1; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDF 3  
111  
DB 5 TDF 7

RESULT 9  
US-08-136-743B-32  
Sequence 32, Application US/08136743B  
Patent No. 5459063  
GENERAL INFORMATION:  
APPLICANT: Barry S. Cooperman, Harvey Rubin,  
Applicant: Jerome Salem, and Allison L. Fisher  
TITLE OF INVENTION: "Plasmodium falciparum Ribonu-  
TITLE OF INVENTION: cleotide Reductase, DNA Sequences Therefor and Peptide Inhi  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The University of Pennsylvania  
STREET: Suite 330  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: U.S.A.  
ZIP: 19104-3246  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 Inch, 720 Kb  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/136,743B  
FILING DATE: 10/14/93  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Monaco, Daniel A.  
REGISTRATION NUMBER: 30,480  
REFERENCE/DOCKET NUMBER: 3957-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-8383  
TELEFAX: (215) 568-5549  
TELEX: NO. 5459063e  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-136-743B-32

Query Match 48.6%; Score 17; DB 1; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDF 3  
111  
DB 5 TDF 7

RESULT 10  
US-08-405-230-5  
Sequence 5, Application US/08405230  
Patent No. 5707846  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
APPLICANT: NISHIKAWA, Atsushi



APPLICANT: YAMAGUCHI, No. 5707846cm1  
TITLE OF INVENTION: NOVEL N-ACETYLGLUCOSAMINYL TRANSFERASE  
TITLE OF INVENTION: GENE CODING THEREFOR AND PROCESS FOR PRODUCTION THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/405,230  
FILING DATE: 16-MAR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/110,736  
FILING DATE: 23-AUG-1993  
APPLICATION NUMBER: JP 4-245950  
FILING DATE: 24-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-237118  
FILING DATE: 06-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 001560-215  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-405-230-5

Query Match 48.6%; Score 17; DB 1; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDF 3  
111  
Db 2 TDF 4

RESULT 11  
US-08-910-990-5  
Sequence 5, Application US/08910990  
Patent No. 5834284  
GENERAL INFORMATION:  
APPLICANT: TANIGUCHI, Naoyuki  
APPLICANT: NISHIKAWA, Atsushi  
APPLICANT: YAMAGUCHI, No. 5834284cm1  
TITLE OF INVENTION: NOVEL N-ACETYLGLUCOSAMINYL TRANSFERASE  
TITLE OF INVENTION: GENE CODING THEREFOR AND PROCESS FOR PRODUCTION THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,990  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,230  
FILING DATE: 16-MAR-1995  
APPLICATION NUMBER: US 08/110,736  
FILING DATE: 23-AUG-1993  
APPLICATION NUMBER: JP 4-245950  
FILING DATE: 24-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-237118  
FILING DATE: 06-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 001560-215  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-910-990-5

Query Match 48.6%; Score 17; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TDF 3  
111  
Db 2 TDF 4

RESULT 12  
US-08-739-401A-6  
Sequence 6, Application US/08739401A  
Patent No. 5837461  
GENERAL INFORMATION:  
APPLICANT: Neitz, Maureen E.  
APPLICANT: Neitz, John F.  
TITLE OF INVENTION: DETECTION OF CONE-PHOTORECEPTOR-BASED  
TITLE OF INVENTION: VISION DISORDERS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 411 East Wisconsin Avenue  
CITY: Milwaukee  
STATE: Wisconsin  
COUNTRY: U.S.A.  
ZIP: 53202-4497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/739,401A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Baker, Jean C.  
REGISTRATION NUMBER: 35,433

REFERENCE/DOCKET NUMBER: 650053.91151  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 277-5709  
TELEFAX: (414) 271-3552  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-739-401A-6

Query Match 48.6%: Score 17; DB 2; Length 7;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 PTLT 6  
111  
Db 2 FTVT 5

RESULT 13  
PCT-US93-11703-72  
Sequence 72, Application PC/TUS9311703  
GENERAL INFORMATION:  
APPLICANT: Chiron Mimotopes Pty. Ltd.  
TITLE OF INVENTION: T-Cell Epitopes  
NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Grant D. Green  
STREET: 4560 Horton St.  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/11703  
FILING DATE: 28-DEC-1993  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/984,852  
FILING DATE: 02-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D.  
REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 0222.101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 510-601-2706  
TELEFAX: 510-655-3542  
INFORMATION FOR SEQ ID NO: 72:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US93-11703-72

Query Match 48.6%: Score 17; DB 5; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DFT 4  
111  
Db 3 DFT 5

RESULT 14  
US-08-180-209B-14  
Sequence 14, Application US/08180209B  
Patent No. 5593877  
GENERAL INFORMATION:  
APPLICANT: King, Te-Piao  
TITLE OF INVENTION: CLONING AND RECOMBINANT PRODUCTION OF  
VESPID VENOM ENZYMES, SUCH AS PHOSPHOLIPASE AND  
HYALURONIDASE, AND IMMUNOLOGICAL THERAPIES BASED  
THEREON  
TITLE OF INVENTION: THEREON  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/180,209B  
FILING DATE: 11-JAN-1994  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/031,400  
FILING DATE: 11-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-074 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-180-209B-14

Query Match 45.7%: Score 16; DB 1; Length 5;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TDLT 4  
111  
Db 2 TDLT 5

RESULT 15  
US-08-385-745-14  
Sequence 14, Application US/08385745  
Patent No. 5612209  
GENERAL INFORMATION:  
APPLICANT: King, Te Piao  
TITLE OF INVENTION: Cloning and Recombinant Production of  
VESPID VENOM PHOSPHOLIPASES, and Immunological Therapies  
THEREON  
TITLE OF INVENTION: Based Thereon  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York

STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/385,745  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/031,400  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Mirock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 3288-020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-385-745-14

Query Match 45.7%; Score 16; DB 1; Length 5;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 0; Mismatches 1; Indels 0;

OY 1 TDPF 4  
|||  
DB 2 TDLF 5

Search completed: July 15, 2002, 13:26:14  
Job time: 494 sec

1.

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:27:09 ; Search time 49.13 seconds  
(without alignments)  
13.691 Million cell updates/sec

Title: US-09-712-819A-6

Perfect score: 33  
Sequence: 1 PTKISR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 455

Minimum DB seq length: 0  
Maximum DB seq length: 7

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	45.5	7	2 S19630	ribosomal protein
2	13	39.4	7	2 A28709	phosphonacetaldeh
3	12	36.4	5	2 T14910	hypothetical prote
4	11	33.3	5	2 PT0644	T-cell receptor be
5	10	30.3	7	2 PS0254	18k protein 5507-
6	9	27.3	3	3 T13892	cytochrome-c oxida
7	9	27.3	4	2 T40804	endoglucanase F -
8	9	27.3	4	2 T46627	hypothetical prote
9	9	27.3	5	2 A60521	glycogen phosphory
10	9	27.3	5	2 E42364	flagellar protein
11	9	27.3	5	2 A44955	alkanal monooxygen
12	9	27.3	5	2 S11127	phosphoprotein, bo
13	9	27.3	5	2 PT0525	T-cell receptor be
14	9	27.3	5	2 PT0577	T-cell receptor be
15	9	27.3	5	2 PT0565	T-cell receptor be
16	9	27.3	5	2 PT0700	T-cell receptor be
17	9	27.3	5	2 S69237	surface protein te
18	9	27.3	6	2 A60986	N-formyl oligopept
19	9	27.3	6	2 A43766	28k ubiquitin-immu
20	9	27.3	6	2 I37263	Y protein - human
21	9	27.3	6	2 B26206	alpha-1,4-glucan-p
22	9	27.3	6	2 I65546	MHC H2-L antigen-
23	9	27.3	6	2 PT0518	T-cell receptor be
24	9	27.3	6	2 PT0662	T-cell receptor be
25	9	27.3	6	2 I49424	cytotoxic T-lympho
26	9	27.3	7	2 JN0859	peptidyl-dipeptida
27	9	27.3	7	2 A15398	cholesterol oxidase (E
28	9	27.3	7	2 B39127	phosphotransferase
29	9	27.3	7	2 S25266	PIIE protein - Esc

30	9	27.3	7	2 A25269	sex pheromone CAM3
31	9	27.3	7	2 A30812	sex pheromone CCF1
32	9	27.3	7	2 PN0649	alpha-dextrin endo
33	9	27.3	7	2 S09066	glutathione IV alpha
34	9	27.3	7	2 PN0150	omega-glutamine 1'
35	9	27.3	7	2 S78024	ribosomal protein
36	9	27.3	7	2 E48394	glycoprotein compo
37	9	27.3	7	2 I48086	DNA topoisomerase
38	9	27.3	7	2 PT0671	T-cell receptor be
39	9	27.3	7	2 S66442	glutathione S-tran
40	9	27.3	7	2 B48394	major fat-globule
41	8	24.2	4	2 I40505	hypothetical prote
42	8	24.2	5	2 I39964	ribosomal protein
43	8	24.2	5	2 I39965	ribosomal protein
44	8	24.2	5	2 B23565	R-phycocyanin al
45	8	24.2	5	2 B23565	

## ALIGNMENTS

## RESULT 1

S19630  
ribosomal protein L30 - Streptomyces griseus (fragment)  
C:Species: Streptomyces griseus  
C:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 06-Jun-1997  
C:Accession: S19630  
R:Ochl, K.  
Int. J. Syst. Bacteriol. 42, 144-150, 1992  
A:Title: Electrophoretic heterogeneity of ribosomal protein Ar-L30 among actinomycete  
A:Reference number: S19630; MUID:92144363  
A:Accession: S19630  
A:Molecule type: protein  
A:Residues: 1-7 <OCH>  
A:Experimental source: strain IFO 13189  
C:Superfamily: Escherichia coli ribosomal protein L30  
C:Keywords: protein biosynthesis; ribosome

Query Match 45.5% Score 15; DB 2; Length 7;  
Best Local Similarity 60.0% Pred. No. 2.8e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LKISR 7  
DB 3 LKITO 7

RESULT 2  
A28709  
phosphonacetaldehyde hydrolase - Bacillus cereus (fragment)  
C:Species: Bacillus cereus  
C:Date: 22-Aug-1988 #sequence\_revision 22-Aug-1988 #text\_change 30-Sep-1993  
C:Accession: A28709  
R:Olsem, D.B.; Hepburn, T.W.; Moos, M.; Mariano, P.S.; Dunaway-Mariano, D.  
Biochemistry 27, 2229-2234, 1988  
A:Title: Investigation of the Bacillus cereus phosphonacetaldehyde hydrolase. Eviden  
A:Reference number: A28709; MUID:86241058  
A:Accession: A28709  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-7 <OLS>

Query Match 39.4% Score 13; DB 2; Length 7;  
Best Local Similarity 100.0% Pred. No. 2.8e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LKI 5  
DB 1 LKI 3

RESULT 3  
T14910  
hypotheical protein - parsley  
C:Species: Petroselinum crispum (parsley)  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: T14910  
R:Kircher, S.; Ledger, S.; Hayashi, H.; Weissnaar, B.; Schafer, E.; Frohnmeyer, H.  
Mol. Gen. Genet. 257, 595-605, 1998  
A:Title: CPRF4, a novel plant bZIP protein of the CPRF family: comparative analysis of  
A:Reference number: 218261; MUID:98265918  
A:Accession: T14910  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-5 <NR>  
A:Cross-references: EMBL:Y10810; NID:g3336904; PIDN:CAA71769.1; PID:g3336905  
A:Experimental source: ssp. Hamburger Schnitt

Query Match 36.4%; Score 12; DB 2; Length 5;  
Best Local Similarity 66.7%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 5 ISR 7  
Db 2 VSR 4

RESULT 4  
PT0644  
T-cell receptor beta chain V-D-J region (111-16) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
C:Accession: PT0644  
R:Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
A:Reference number: PT0509; MUID:91277601  
A:Accession: PT0644  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-5 <PEE>  
A:Experimental source: newborn thymus, strain BALB/c  
C:Keywords: T-cell receptor

Query Match 33.3%; Score 11; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FT 2  
Db 4 FT 5

RESULT 5  
PS0254  
18K protein 5507 - rice (strain Nihonbare) (fragment)  
C:Species: Oryza sativa (rice)  
C>Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 23-Mar-1995  
C:Accession: PS0254  
R:Tsugita, A.  
submitted to JIPID, April 1993  
A:Reference number: PS0206  
A:Accession: PS0254  
A:Molecule type: protein  
A:Residues: 1-7 <TSID>  
A:Experimental source: leaf, chloroplast, strain Nihonbare  
A:Note: Molecular weight 18K, pI 4.4

Query Match 30.3%; Score 10; DB 2; Length 7;  
Best Local Similarity 40.0%; Pred. No. 2.8e+05;

Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 LKISR 7  
Db 1 LAIAK 5

RESULT 6  
T13892  
cytochrome-c oxidase (EC 1.9.3.1) chain I [Imported] - river lamprey mitochondrion (f  
C:Species: mitochondrion lampetra fluviatilis (river lamprey)  
C>Date: 15-Jun-2001 #sequence\_revision 15-Jun-2001 #text\_change 15-Jun-2001  
C:Accession: T13892  
R:Delarbre, C.; Bartel, V.; Tiller, S.; Janvier, P.; Gachelin, G.  
Mol. Biol. Evol. 14, 807-813, 1997  
A:Title: The main features of the crnulate mitochondrial DNA between the NDI and the  
A:Reference number: 217775; MUID:97398704  
A:Accession: T13892  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3 <DEL>  
A:Cross-references: EMBL:Y09528; NID:g2340016; PIDN:CAA70721.1; PID:g4379123  
C:Genetics:  
A:Genome: mitochondrion  
A:Note: COI  
C:Keywords: mitochondrion; oxidoreductase

Query Match 27.3%; Score 9; DB 3; Length 3;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TL 3  
Db 2 TL 3

RESULT 7  
I40804  
endoglucanase F - Clostridium thermocellum (fragment)  
C:Species: Clostridium thermocellum  
C>Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 16-Aug-1996  
C:Accession: I40804  
R:Mishra, S.; Beguin, P.; Aubert, J.  
J. Bacteriol. 173, 80-85, 1991  
A:Title: Transcription of clostridium thermocellum endoglucanase genes celf and celd  
A:Reference number: I40804; MUID:91100322  
A:Accession: I40804  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-4 <RES>  
A:Cross-references: GB:M64363; NID:g144771  
C:Genetics:  
A:Gene: celf  
A:Start codon: TTG

Query Match 27.3%; Score 9; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KI 5  
Db 3 KI 4

RESULT 8  
T46627  
hypotheical protein c4 - loblolly pine  
C:Species: Pinus taeda (loblolly pine)  
C>Date: 18-Feb-2000 #sequence\_revision 18-Feb-2000 #text\_change 18-Feb-2000  
C:Accession: T46627  
R:Chang, S.; Puryear, J.; Funkhouser, E.A.; Newton, R.J.; Cairney, J.

submitted to the EMBL Data Library, July 1995  
A:Description: Cloning of a chitinase homolog which lacks chitin binding sites and is de  
A:Reference number: Z23105  
A:Accession: T46627  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-4 <CH>  
A:Cross-references: EMBL:U31309; NID:9974285; PID:9974292  
A:Experimental source: strain sopTxs6P3; 8 month seedlings

Query Match 27.3%; Score 9; DB 2; Length 4;  
Best Local Similarity 33.3%; Pred. No. 2.8e+05;  
Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 LKI 5  
DB 1 MKL 3

RESULT 9  
A60521  
glycogen phosphorylase (EC 2.4.1.1), muscle - mullet (Liza ramada) (fragment)  
N:Alternate names: glycogen phosphorylase b  
C:Species: Liza ramada  
C>Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 11-May-2000  
C:Accession: A60521  
R:Bonomusa, L.; Paanante, I.V.  
Comp. Biochem. Physiol. B 95, 295-301, 1990  
A:Title: Purification and characterization of glycogen phosphorylase B from skeletal mus  
A:Reference number: A60521; MUID:90227907  
A:Accession: A60521  
A:Molecule type: protein  
A:Residues: 1-5 <BN>  
A:Superfamily: phosphorylase  
C:Keywords: glycyltransferase; hexosyltransferase; phosphoprotein  
F3/Binding site: phosphate (Ser) (covalent) (by phosphorylase b kinase) #status experin

Query Match 27.3%; Score 9; DB 2; Length 5;  
Best Local Similarity 66.7%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 KIS 6  
DB 1 QIS 3

RESULT 10  
E42364  
flagellar protein fljR - Salmonella typhimurium (fragment)  
C:Species: Salmonella typhimurium  
C>Date: 24-Jul-1992 #sequence\_revision 24-Jul-1992 #text\_change 30-Sep-1993  
C:Accession: E42364  
R:Voigter, A.P.; Homma, M.; Irikura, V.M.; Macnab, R.M.  
J. Bacteriol. 173, 3564-3572, 1991  
A:Title: Salmonella typhimurium mutants defective in flagellar filament regrowth and seq  
A:Reference number: A42364; MUID:91258342  
A:Accession: E42364  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-5 <VO>  
A:Cross-references: GB:M62408

Query Match 27.3%; Score 9; DB 2; Length 5;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TL 3  
DB 3 TL 4

RESULT 11  
A44955  
alkanal monooxygenase (FMN-linked) (EC 1.14.14.3) alpha chain - Vibrio harveyi (frag  
C:Species: Vibrio harveyi  
C>Date: 03-Jun-1993 #sequence\_revision 03-Jun-1993 #text\_change 26-May-2000  
C:Accession: A44955  
R:Paquette, O.; Tu, S.C.  
Photochem. Photobiol. 50, 817-825, 1989  
A:Title: Chemical modification and characterization of the alpha cysteine 106 at the  
A:Reference number: A44955; MUID:90175700  
A:Accession: A44955  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-5 <PA>  
C:Keywords: FMN; luminescence; monooxygenase; oxidoreductase

Query Match 27.3%; Score 9; DB 2; Length 5;  
Best Local Similarity 66.7%; Pred. No. 2.8e+05;  
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 ISR 7  
DB 3 IXR 5

RESULT 12  
S11127  
phosphoprotein, bone - chicken (fragment)  
C:Species: Gallus gallus (chicken)  
C>Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 04-Mar-2000  
C:Accession: S11127; S11128  
R:Mikuni-Takagaki, Y.; Gilmcher, M.J.  
Biochem. J. 268, 585-591, 1990  
A:Title: Post-translational processing of chicken bone phosphoproteins. Identification  
A:Reference number: S11127; MUID:90303246  
A:Accession: S11127  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-5 <MI>  
A:Accession: S11128  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 'X', 2-5 <MIK2>  
C:Keywords: phosphoprotein

Query Match 27.3%; Score 9; DB 2; Length 5;  
Best Local Similarity 33.3%; Pred. No. 2.8e+05;  
Matches 1; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 5 ISR 7  
DB 3 VSK 5

RESULT 13  
PT0525  
T-cell receptor beta chain V-D-J region (100-41) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
C:Accession: PT0525  
R:Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A:Title: Unclonal sequences of fetal T cell receptor beta chains have few N regions  
A:Reference number: PT0525; MUID:91277601  
A:Accession: PT0525  
A>Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-5 <FE>  
A:Experimental source: adult thymus, strain BALB/c  
C:Keywords: T-cell receptor

Query Match 27.3%; Score 9; DB 2; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 SR 7  
 ||  
 Db 2 SR 3

## RESULT 14

PT0577  
 T-cell receptor beta chain V-D-J region (141-1BC) - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
 C:Accession: PT0577; PT0574  
 R:Feeney, A.J.  
 J. Exp. Med. 174, 115-124, 1991  
 A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
 A:Reference number: PT0509; MUID:91277601  
 A:Accession: PT0577  
 A>Status: translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-5 <PEP>  
 A:Experimental source: day 19 fetal thymus, strain BALB/c, clone 141-1BC  
 A:Accession: PT0574  
 A:Molecule type: translation not shown  
 A:Residues: 1-5 <PEP>  
 A:Experimental source: day 19 fetal thymus, strain BALB/c, clone 141-1Q  
 C:Keywords: T-cell receptor

Query Match 27.3%; Score 9; DB 2; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 SR 7  
 ||  
 Db 3 SR 4

## RESULT 15

PT0565  
 T-cell receptor beta chain V-D-J region (141-1CF) - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
 C:Accession: PT0565  
 R:Feeney, A.J.  
 J. Exp. Med. 174, 115-124, 1991  
 A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
 A:Reference number: PT0509; MUID:91277601  
 A:Accession: PT0565  
 A>Status: translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-5 <PEP>  
 A:Experimental source: day 19 fetal thymus, strain BALB/c  
 C:Keywords: T-cell receptor

Query Match 27.3%; Score 9; DB 2; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 SR 7  
 ||  
 Db 3 SR 4

Search completed: July 15, 2002, 13:27:09  
 Job time: 439 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:38:10 ; Search time 24.49 seconds  
(without alignments)  
11.067 Million cell updates/sec

Title: US-09-712-819A-6

Perfect score: 33

Sequence: 1 FTKISR 7

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 84

Minimum DB seq length: 0

Maximum DB seq length: 7

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt-40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	11	33.3	5 1	RE31_LITRU
2	11	33.3	5 1	RE32_LITRU
3	10	30.3	7 1	GFRP_MOUSE
4	9	27.3	6 1	UF01_MOUSE
5	9	27.3	6 1	UN06_CLOPA
6	9	27.3	7 1	CCPI_EMMPA
7	9	27.3	7 1	CHOX_ALCSP
8	9	27.3	7 1	CIA_ENTFA
9	9	27.3	7 1	UC24_MAIZE
10	8	24.2	5 1	AL14_CARMA
11	8	24.2	5 1	PSK_DAUCA
12	8	24.2	7 1	AL12_CARMA
13	8	24.2	7 1	AL13_CARMA
14	8	24.2	7 1	AL14_CARMA
15	8	24.2	7 1	AL15_CARMA
16	8	24.2	7 1	AL17_CTDPO
17	8	24.2	7 1	CARF_MYTED
18	8	24.2	7 1	CARE_MYTED
19	7	21.2	3 1	LUXE_VIBFI
20	7	21.2	6 1	LOKI_LOCOMI
21	7	21.2	6 1	VP19_HSVIK
22	6	18.2	4 1	ACHT_ACHPT
23	6	18.2	4 1	PAR3_HIRME
24	6	18.2	4 1	PAR4_HIRME
25	6	18.2	4 1	FFKA_ANTEL
26	6	18.2	4 1	FLRF_HIRME
27	6	18.2	4 1	FLRN_ANTEL
28	6	18.2	4 1	FMRE_MACNI
29	6	18.2	4 1	FYRI_ANTEL
30	6	18.2	4 1	OCPI_OCTMI
31	6	18.2	5 1	ET03_LITRU
32	6	18.2	5 1	ET04_LITRU
33	6	18.2	5 1	FARP_ANTIR

34	6	18.2	5 1	PAP2_PARMA	P81864 pardachirus
35	6	18.2	5 1	RE11_LITRU	P82070 litorea rub
36	6	18.2	5 1	RE21_LITRU	P82071 litorea rub
37	6	18.2	5 1	SUGA_ACHDO	P19991 acheta dome
38	6	18.2	5 1	TP15_CANFA	P54714 canis fam1
39	6	18.2	5 1	UC22_MAIZE	P80628 zea mays (m
40	6	18.2	6 1	ACPH_RABIT	P25154 oryctolagus
41	6	18.2	6 1	CIP1_MYTED	P13736 mytilus edu
42	6	18.2	6 1	CIP2_MYTED	P13737 mytilus edu
43	6	18.2	6 1	ET01_LITRU	P82096 litorea rub
44	6	18.2	6 1	FARP_MONEX	P41966 moniezia ex
45	6	18.2	7 1	ET05_LITRU	P82101 litorea rub

## ALIGNMENTS

RESULT 1					ALIGNMENTS				
ID	RE31_LITRU	STANDARD:	PRT:	5 AA.	ID	RE31_LITRU	STANDARD:	PRT:	5 AA.
AC	P82072;				AC	P82072;			
DT	01-MAR-2002 (Rel. 41, Created)				DT	01-MAR-2002 (Rel. 41, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)				DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Rubellidin 3.1.				DE	Rubellidin 3.1.			
OS	Litorea rubella (Desert tree frog).				OS	Litorea rubella (Desert tree frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;				OC	Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;			
OC	Litorea.				OC	Litorea.			
CX	NCBI_TaxID-104895;				CX	NCBI_TaxID-104895;			
RP	[1]				RP	[1]			
RM	SEQUENCE, AND MASS SPECTROMETRY.				RM	SEQUENCE, AND MASS SPECTROMETRY.			
RC	TISSUE-Skin secretion;				RC	TISSUE-Skin secretion;			
RA	Steinborner S.T., Wabnitz P.A., Waugh R.J., Bowle J.H., Gao C.,				RA	Steinborner S.T., Wabnitz P.A., Waugh R.J., Bowle J.H., Gao C.,			
RT	Tyler M.J., Wallace J.C.;				RT	Tyler M.J., Wallace J.C.;			
RT	"The structure of new peptides from the Australian red tree frog				RT	"The structure of new peptides from the Australian red tree frog			
RT	'Litorea rubella', the skin peptide profile as a probe for the study				RT	'Litorea rubella', the skin peptide profile as a probe for the study			
RT	of evolutionary trends of amphibians.";				RT	of evolutionary trends of amphibians.";			
RL	Aust. J. Chem. 49:955-963(1996).				RL	Aust. J. Chem. 49:955-963(1996).			
CC	-1- FUNCTION: SHOW NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC				CC	-1- FUNCTION: SHOW NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC			
CC	ACTIVITY.				CC	ACTIVITY.			
CC	-1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.				CC	-1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.			
CC	-1- MASS SPECTROMETRY: MW-655; METHOD=FAV.				CC	-1- MASS SPECTROMETRY: MW-655; METHOD=FAV.			
KW	Amphibian skin; Amidation.				KW	Amphibian skin; Amidation.			
FT	MOD_RES 5				FT	MOD_RES 5			
FT	SEQUENCE 5 AA: 656 MW: 71A9C9CB10300000 CRC64;				FT	SEQUENCE 5 AA: 656 MW: 71A9C9CB10300000 CRC64;			
Query Match 33.3%; Score 11; DB 1; Length 5;									
Best Local Similarity 100.0%; Pred. No. 1e+05;									
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
OY	1 FT 2				OY	1 FT 2			
DB	4 FT 5				DB	4 FT 5			
RESULT 2									
ID	RE32_LITRU	STANDARD:	PRT:	5 AA.	ID	RE32_LITRU	STANDARD:	PRT:	5 AA.
AC	P82073;				AC	P82073;			
DT	01-MAR-2002 (Rel. 41, Created)				DT	01-MAR-2002 (Rel. 41, Created)			
DT	01-MAR-2002 (Rel. 41, Last sequence update)				DT	01-MAR-2002 (Rel. 41, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)				DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Rubellidin 3.2				DE	Rubellidin 3.2			
OS	Litorea rubella (Desert tree frog).				OS	Litorea rubella (Desert tree frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;				OC	Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;			
OC	Litorea.				OC	Litorea.			
OX	NCBI_TaxID-104895;				OX	NCBI_TaxID-104895;			
RN	[1]				RN	[1]			
RP	SEQUENCE.				RP	SEQUENCE.			
RC	TISSUE-Skin secretion;				RC	TISSUE-Skin secretion;			

RA Mabiniz P.A., Bowie J.H., Tyler M.J., Wallace J.C.:  
 "Peptides from the skin glands of the Australian buzzing tree frog  
 RT *Litoria electrica*. Comparison with the skin peptides from *Litoria*  
 RT *rubella*".  
 RL Aust. J. Chem. 52:0-0(1999).  
 CC -1- FUNCTION: SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR ANTI-BIOTIC  
 CC ACTIVITY.  
 CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
 KW Amphibian skin.  
 SQ SEQUENCE 5 AA; 570 MW; 71A9C9C862A00000 CRC64;

Query Match 33.3%; Score 11; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FT 2  
 1 1  
 4 FT 5

RESULT 3  
 GERP\_MOUSE  
 ID GERP\_MOUSE STANDARD; PRT; 7 AA.  
 AC P99025;  
 DT 13-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-OCT-2001 (Rel. 40, Last annotation update)  
 DE GTP cyclohydrolase I feedback regulatory protein (P35) (Fragment).  
 GN GCHFR OR GFRP.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE-LIVER;  
 RA Sanchez J.-C., Rouge V., Frutiger S., Hughes G., Van J.X.,  
 RA Hoogland C., Appel R.D., Binz P.-A., Hochstrasser D.F.,  
 RA Cowthorne M.;  
 RL Submitted (AUG-1998) to the SWISS-PROT data bank.  
 CC -1- FUNCTION: MEDIATES TETRAHYDROBIOPERIN INHIBITION OF GTP  
 CC CYCLOHYDROLASE I. THIS INHIBITION IS REVERSED BY L-PHENYLANILINE  
 CC (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC DR SWISS-2DPAGE; P99025; MOUSE.  
 FT INIT\_MET 0  
 FT NON\_TER 7  
 SQ SEQUENCE 7 AA; 806 MW; 71B5B057273B4700 CRC64;

Query Match 30.3%; Score 10; DB 1; Length 7;  
 Best Local Similarity 75.0%; Pred. No. 1e+05;  
 Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 LKTS 6  
 1 1 1  
 3 LKTS 6

RESULT 4  
 UFI1\_MOUSE  
 ID UFI1\_MOUSE STANDARD; PRT; 5 AA.  
 AC P38639;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)  
 DE Unknown protein from 2D-page of fibroblasts (P19) (Fragment).  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]

RP SEQUENCE.  
 RC TISSUE-Fibroblast;  
 RC MEDLINE=95009907; PubMed=7523108;  
 RA Merrick B.A., Patterson R.M., Wichter L.L., He C., Solkirk J.R.;  
 RT "Separation and sequencing of familial and novel murine proteins  
 RT using preparative two-dimensional gel electrophoresis".  
 RL Electrophoresis 15:735-743(1994).  
 CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN  
 CC PROTEIN IS: 6.6, ITS MW IS: 19 kDa.  
 FT NON\_TER 5  
 FT SEQUENCE 5 AA; 717 MW; 7364087043100000 CRC64;

Query Match 27.3%; Score 9; DB 1; Length 5;  
 Best Local Similarity 66.7%; Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 ISR 7  
 1 1  
 2 IGR 4

RESULT 5  
 UN06\_CLOPA  
 ID UN06\_CLOPA STANDARD; PRT; 6 AA.  
 AC P81351;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Unknown protein CP 6 from 2D-page (Fragment).  
 OS Clostridium pasteurianum.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 CC Clostridium.  
 OX NCBI\_TaxID=1501;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN-W5;  
 RA MEDLINE=96291870; PubMed=9629918;  
 RA Flensburg R., Skjeldal L.;  
 RT "Two-dimensional gel electrophoresis separation and N-terminal  
 RT sequence analysis of proteins from Clostridium pasteurianum W5";  
 RL Electrophoresis 19:802-806(1998).  
 CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN  
 CC PROTEIN IS: 5.0, ITS MW IS: 75.9 kDa.  
 FT NON\_TER 6  
 FT SEQUENCE 6 AA; 657 MW; 605B1DC1445A8000 CRC64;

Query Match 27.3%; Score 9; DB 1; Length 6;  
 Best Local Similarity 50.0%; Pred. No. 1e+05;  
 Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 LKTI 5  
 1 1  
 3 TAEI 6

RESULT 6  
 CCE1\_ENTPA  
 ID CCE1\_ENTPA STANDARD; PRT; 7 AA.  
 AC P20104;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-FEB-1991 (Rel. 17, Last annotation update)  
 DE Sex pheromone CCF10.  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;  
 CC Enterococcus.  
 OX NCBI\_TaxID=1351;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=89008313; PubMed=3139658;  
 RA Mori M., Sakagami Y., Ishii Y., Isogai A., Kitada C., Fujino M.,

```

RA Adsit J.C., Dunny G.M., Suzuki A.;
RT *Structure of CCP10, a peptide sex pheromone which induces
RT conjugative transfer of the Streptococcus faecalis tetracycline
RT resistance plasmid, pCF10.*;
RL J. Biol. Chem. 263:14574-14578(1988).
CC -1- FUNCTION: CCP10 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
CC HEMOLYSIN PLASMID PCF10.
DR PIR: A30812; A30812.
KW Pheromone.
SQ SEQUENCE 7 AA: 72C9D2C731B2C740 CRC64;

Query Match 27.3%; Score 9; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TL 3
   1 1
Db 3 TL 4

RESULT 7
CHOX_ALCSP STANDARD; PRT; 7 AA.
ID CHOX_ALCSP
AC P16101;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last annotation update)
DE Choline oxidase (EC 1.1.3.17) (Fragment).
OS Alcaligenes sp.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Alcaligenes.
OX NCBI_TaxID=512;
RN [1]
RP SEQUENCE.
RX MEDLINE=81006769; PubMed=6997283;
RA Ohta-Fukuyama M., Miyake Y., Emi S., Yamano T.;
RT *Identification and properties of the prosthetic group of choline
RT oxidase from Alcaligenes sp.*;
RL J. Biochem. 88:197-203(1980).
CC -1- CATALYTIC ACTIVITY: Choline + O(2) = betaine aldehyde + H(2)O(2).
DR PIR: A15398; A15398.
KW Oxidoreductase.
FT NON_TER
SQ SEQUENCE 7 AA: 839 MW; 7415B1B457644AC0 CRC64;

Query Match 27.3%; Score 9; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 SR 7
   1 1
Db 6 SR 7

RESULT 8
CIA_ENTFA STANDARD; PRT; 7 AA.
ID CIA_ENTFA
AC P11932;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Sex pheromone CAM373 (Clumping-inducing agent) (CIA).
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE.
RX MEDLINE=87005252; PubMed=3093276;
RA Moti M., Tanaka H., Sakagami Y., Isogai A., Fujino M., Kitada C.,
RA White B.A., An F.Y., Clewell D.B., Suzuki A.;

```

```

RT *Isolation and structure of the Streptococcus faecalis sex pheromone,
RT CAM373.*;
RL FEBS Lett. 206:69-72(1986).
CC -1- FUNCTION: CAM373 INDUCES MATING RESPONSE OF DONOR CELLS
CC HARBORING PAM373.
CC -1- MISCELLANEOUS: THE N-TERMINUS IS POSSIBLY RESPONSIBLE FOR
CC SPECIFICITY OF PHEROMONES TO PLASMIDS.
CC -1- SIMILARITY: C-TERMINAL TO THE PHEROMONES CPD1 AND CAD1.
DR PIR: A25269; A25269.
KW Pheromone.
SQ SEQUENCE 7 AA: 734 MW; 75BDD72059C05DB0 CRC64;

Query Match 27.3%; Score 9; DB 1; Length 7;
Best Local Similarity 66.7%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTL 3
   1 1
Db 3 FTL 5

RESULT 9
UC24_MAIZE STANDARD; PRT; 7 AA.
ID UC24_MAIZE
AC F80630;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Unknown protein from 2D-page of etiolated coleoptile (Spot 447)
DE (Fragment).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE.
RX TISSUE=coleoptile;
RA Toulzet P., Riccardi F., Morin C., Damerval C., Huot J.-C.,
RA Fernoulet J.-C., Zivy M., de Vienne D.;
RT *The maize two dimensional gel protein database: towards an integrated
RT genome analysis program.*;
RL Theor. Appl. Genet. 93:997-1005(1996).
CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 6.0, ITS MW IS: 30.0 kDa.
DR MaizeDB: P80630; COLEOPTILE.
FT NON_TER
SQ SEQUENCE 7 AA: 665 MW; 60C1B5B33DC1B5D0 CRC64;

Query Match 27.3%; Score 9; DB 1; Length 7;
Best Local Similarity 66.7%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 TLK 4
   1 1
Db 2 TLK 4

RESULT 10
AL14_CARMA STANDARD; PRT; 5 AA.
ID AL14_CARMA
AC P81817;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Carcinustatin 14.
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;

```

OC Eubrachyura: Portunoidae; Portunidae; Carcinus.  
OX NCBI\_Taxid=6759;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
RX MEDLINE=98121193; PubMed=9461295;  
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
Thorpé A.;  
RT "Isolation and identification of multiple neuropeptides of the  
allatostatin superfamily in the shore crab *Carcinus maenas*.";  
RL Eur. J. Biochem. 250:727-734(1997).  
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
KM Neuropeptide; BELONGS TO THE ALLATOSTATIN FAMILY.  
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
KW Neuropeptide; Amidation; Multigene family.  
FT MOD\_RES 5  
SQ SEQUENCE 5 AA; 586 MW; 672879D5AB300000 CRC64;

Query Match 24.2%; Score 8; DB 1; Length 5;  
Best Local Similarity 66.7%; Pred. No. 1e+05;  
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTL 3  
DB 3 FGL 5

## RESULT 11

PSK\_DAUCA STANDARD; PRT; 5 AA.  
ID PSK\_DAUCA  
AC P86261;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Phytosulfokine-alpha (PSK-alpha) [contains: Phytosulfokine-beta (PSK-beta)].  
OS Daucus carota (Carrot).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
CC Asteridae; eusterids II; Apiales; Apiaceae; Daucus.  
OX NCBI\_Taxid=4039;  
RN [1]  
RP SEQUENCE, AND IDENTIFICATION BY MASS SPECTROMETRY.  
RC STRAIN=cv. US-Harumakigosun;  
RX MEDLINE=20212743; PubMed=10750705;  
RA Hanei H., Matsuno T., Yamamoto M., Matsubayashi Y., Kobayashi T.,  
Komada H., Sakagami Y.;  
RT "A secreted peptide growth factor, phytosulfokine, acting as a  
stimulatory factor of carrot somatic embryo formation.";  
RL Plant Cell Physiol. 41:27-32(2000).  
CC -1- FUNCTION: IN PRESENCE OF 2,4-D, STIMULATES PROLIFERATION OF THE  
CELLS, BUT DOES NOT STIMULATE DIFFERENTIATION INTO THE SOMATIC  
EMBRYOS.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- PTM: SULFATION IS IMPORTANT FOR ACTIVITY AND FOR THE BINDING TO A  
PUTATIVE MEMBRANE RECEPTOR (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE PHYTOSULFOKINE FAMILY.  
KM Growth factor; Sulfation.  
FT PEPTIDE 1 4  
FT MOD\_RES 1 1  
FT MOD\_RES 3 3  
SQ SEQUENCE 5 AA; 687 MW; 76C1BB504B300000 CRC64;

Query Match 24.2%; Score 8; DB 1; Length 5;  
Best Local Similarity 50.0%; Pred. No. 1e+05;  
Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTL 2  
DB 3 YL 4

RESULT 12  
ALL2\_CARMA STANDARD; PRT; 7 AA.  
ID ALL2\_CARMA  
AC P81805;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Carcinustatin 2.  
OS Carcinus maenas (Common shore crab) (Green crab).  
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
CC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
OX NCBI\_Taxid=6759;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
RX MEDLINE=98121193; PubMed=9461295;  
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
Thorpé A.;  
RT "Isolation and identification of multiple neuropeptides of the  
allatostatin superfamily in the shore crab *Carcinus maenas*.";  
RL Eur. J. Biochem. 250:727-734(1997).  
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
KM Neuropeptide; Amidation; Multigene family.  
FT MOD\_RES 7  
SQ SEQUENCE 7 AA; 770 MW; 672879CDBC85DB70 CRC64;

Query Match 24.2%; Score 8; DB 1; Length 7;  
Best Local Similarity 66.7%; Pred. No. 1e+05;  
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTL 3  
DB 5 FGL 7

RESULT 13  
ALL3\_CARMA STANDARD; PRT; 7 AA.  
ID ALL3\_CARMA  
AC P81806;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Carcinustatin 3.  
OS Carcinus maenas (Common shore crab) (Green crab).  
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
CC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
OX NCBI\_Taxid=6759;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
RX MEDLINE=98121193; PubMed=9461295;  
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
Thorpé A.;  
RT "Isolation and identification of multiple neuropeptides of the  
allatostatin superfamily in the shore crab *Carcinus maenas*.";  
RL Eur. J. Biochem. 250:727-734(1997).  
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
KM Neuropeptide; Multigene family.  
FT MOD\_RES 7  
SQ SEQUENCE 7 AA; 796 MW; 672879CDBC8476B70 CRC64;

Query Match 24.2%; Score 8; DB 1; Length 7;  
Best Local Similarity 66.7%; Pred. No. 1e+05;  
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTL 3  
DB 5 FGL 7

Db 1 1  
5 FGL 7

Search completed: July 15, 2002, 13:38:11  
Job time: 711 sec

RESULT 14  
ALL4\_CARMA STANDARD; PRT; 7 AA.  
AC P81807;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE Carcinustatin 4. (Common shore crab) (Green crab).  
OS Carcinus maenas (Common shore crab) (Green crab).  
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
OX NCBI\_TaxID=6759;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
RX MEDLINE=98121193; PubMed=9461295;  
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
Thorpe A.;  
RT Isolation and identification of multiple neuropeptides of the  
allatostatin superfamily in the shore crab Carcinus maenas.";  
RL Eur. J. Biochem. 250:727-734(1997).  
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
KW Neuropeptide; Multigene family.  
SQ SEQUENCE 7 AA; 782 MW; 672879CDBC8476AC0 CRC64;

Query Match 24.2%; Score 8; DB 1; Length 7;  
Best Local Similarity 66.7%; Pred. No. 1e+05;  
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTL 3  
1 1  
5 FGL 7

RESULT 15  
ALL5\_CARMA STANDARD; PRT; 7 AA.  
ID ALL5\_CARMA  
AC P81808;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE Carcinustatin 5.  
OS Carcinus maenas (Common shore crab) (Green crab).  
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
OX NCBI\_TaxID=6759;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
RX MEDLINE=98121193; PubMed=9461295;  
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
Thorpe A.;  
RT Isolation and identification of multiple neuropeptides of the  
allatostatin superfamily in the shore crab Carcinus maenas.";  
RL Eur. J. Biochem. 250:727-734(1997).  
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
KW Neuropeptide; Multigene family.  
FT MOD\_RES 7  
SQ SEQUENCE 7 AA; 781 MW; 672879CDBC8476420 CRC64;

Query Match 24.2%; Score 8; DB 1; Length 7;  
Best Local Similarity 66.7%; Pred. No. 1e+05;  
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTL 3

10x10  
CHAIN

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd..

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:37:39 ; Search time 86.73 Seconds  
(without alignments)  
13.962 Million cell updates/sec

Title: US-09-712-819A-6  
Perfect score: 33  
Sequence: 1 FTLKISR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 65

Minimum DB seq length: 0  
Maximum DB seq length: 7

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP.EMBRL\_19:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.mhc:\*  
9: sp.organelle:\*  
10: sp.phage:\*  
11: sp.plant:\*  
12: sp.rodent:\*  
13: sp.virus:\*  
14: sp.vertibrate:\*  
15: sp.unclassified:\*  
16: sp.virus:\*  
17: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	39.4	6	10	P82181 spinacia ol
2	13	39.4	6	10	P82182 spinacia ol
3	11	33.3	5	13	P82072 litoria rub
4	11	33.3	5	13	P82073 litoria rub
5	11	33.3	7	8	Q95945 saccharomyc
6	10	30.3	6	10	P82541 spinacia ol
7	9	27.3	4	11	Q08433 rattus norv
8	9	27.3	7	2	Q07354
9	9	27.3	7	4	Q15897 homo sapien
10	9	27.3	7	10	P93233 lycopersico
11	9	27.3	7	11	Q63480 rattus norv
12	9	27.3	7	11	Q55184 rattus norv
13	8	24.2	7	2	Q47029 enterobacte
14	8	24.2	7	2	P70804 azotobacter
15	8	24.2	7	12	Q9Y1R0 human adeno
16	8	24.2	7	12	Q9Y1Q9 human adeno

17	8	24.2	7	12	Q9YVE3	Q9YVE3 human adeno
18	7	21.2	5	2	P83073	P83073 bacillus ce
19	7	21.2	7	2	P72081	P72081 nocardia la
20	7	21.2	7	12	Q66205	Q66205 transmissib
21	7	21.2	7	12	Q9Y010	Q9Y010 transmissib
22	6	18.2	5	13	P82070	P82070 litoria rub
23	6	18.2	5	13	P82071	P82071 litoria rub
24	6	18.2	5	13	P82099	P82099 litoria rub
25	6	18.2	5	13	P82100	P82100 litoria rub
26	6	18.2	6	13	P82096	P82096 litoria rub
27	6	18.2	7	2	Q50556	Q50556 actinobacill
28	6	18.2	7	2	Q34028	Q34028 sphingomona
29	6	18.2	7	2	Q54248	Q54248 streptomyce
30	6	18.2	7	4	Q15903	Q15903 homo sapien
31	6	18.2	7	6	Q28742	Q28742 oryctolagus
32	6	18.2	7	8	P92421	P92421 palthyrosta
33	6	18.2	7	8	P92385	P92385 hordeum mar
34	6	18.2	7	8	P92372	P92372 haynaldia v
35	6	18.2	7	8	P92403	P92403 lophopyrum
36	6	18.2	7	8	P92405	P92405 pseudoroegn
37	6	18.2	7	8	P92387	P92387 henrardla p
38	6	18.2	7	8	P92427	P92427 peridictyon
39	6	18.2	7	8	P92390	P92390 heteranthe
40	6	18.2	7	8	P92226	P92226 crithopsis
41	6	18.2	7	8	P92214	P92214 amblyopyrum
42	6	18.2	7	8	P92430	P92430 aegilops ta
43	6	18.2	7	8	P92221	P92221 bromus iner
44	6	18.2	7	8	P92442	P92442 taeniathe
45	6	18.2	7	8	P92381	P92381 hordeum bra

## ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	AA.
P82181	P82181			
AC	P82181			
DT	01-JUN-2000 (T.EMBLrel. 14, Created)			
DT	01-JUN-2000 (T.EMBLrel. 14, Last sequence update)			
DT	01-JUN-2001 (T.EMBLrel. 17, Last annotation update)			
DE	CHLOROPLAST 50S RIBOSOMAL PROTEIN L10 BETA (FRAGMENT).			
OS	Spinacia oleracea (Spinach).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
OC	Caryophyllales; Caryophyllales; Chenopodiaceae; Spinacia.			
OX	NCBI_TaxID=3562;			
RM	[1]			
RP	SEQUENCE.			
RC	STRAIN=CV, ALMARO; TISSUE=LEAF.			
RX	MEDLINE=20435798; PubMed=10874046;			
RA	Yamaguchi K., Subramanian A.R.;			
RT	"The plastid ribosomal proteins. Identification of all the proteins in			
RT	the 50 S subunit of an organelle ribosome (chloroplast).";			
RL	J. Biol. Chem. 275:28466-28482(2000).			
-1-	FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.			
-1-	SUBCELLULAR LOCATION: CHLOROPLAST.			
-1-	TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES.			
CC	-1- MISCELLANEOUS: ON THE 2D-GEL ITS MW IS: 16.5 KDA.			
CC	-1- SIMILARITY: BELONGS TO THE L10 FAMILY OF RIBOSOMAL PROTEINS.			
DR	InterPro: IPR001790; Ribosomal_L10.			
DR	InterPro: IPR002363; Ribosomal_L10_eub.			
DR	Pfam: PF00466; Ribosomal_L10; PARTIAL.			
DR	Prosite: PS01109; RIBOSOMAL_L10; PARTIAL.			
KW	Ribosomal protein; Chloroplast; rRNA-binding.			
FT	NON TER			
SQ	SEQUENCE 6 AA; 675 MW; 632B415B05DB000 CRC64;			

Query Match 39.4% Score 13; DB 10; Length 6;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ISR 7  
111  
Db 2 ISR 4

RESULT 2

P82182

PRELIMINARY; PRT; 6 AA.

ID P82182;  
AC P82182;  
DT 01-JUN-2000 (TREMBLrel. 14, Created)  
DT 01-JUN-2000 (TREMBLrel. 14, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
OS CHLOROPLAST 50S RIBOSOMAL PROTEIN L10 GAMMA (FRAGMENT).  
OS Spinacia oleracea (Spinach).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
OX NCBI\_TaxID=3562;  
RN [1]  
RP SEQUENCE.

RC STRAIN=CV. ALHARO; TISSUE=LEAF.  
RX MEDLINE=20435796; PubMed=10874046;  
RA Yamaguchi K., Subramanian A.R.;  
RT "The plastid ribosomal proteins. Identification of all the proteins in  
the 50 S subunit of an organelle ribosome (chloroplast)."  
RL J. Biol. Chem. 275:28466-28482(2000).  
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.  
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES.  
CC -1- MISCELLANEOUS: ON THE 2D-GEL ITS MW IS: 16.5 KDa.  
DR InterPro: IPR001790; RIBOSOMAL\_L10.  
DR InterPro: IPR002363; RIBOSOMAL\_L10\_eub.  
DR Pfam: PF00466; RIBOSOMAL\_L10; PARTIAL.  
DR PROSITE: PS01109; RIBOSOMAL\_L10; PARTIAL.  
KW RIBOSOMAL protein; chloroplast; RNA-binding.  
FT NON\_TER  
SQ SEQUENCE 6 AA: 675 MW: 63218415805DB000 CRC64;

Query Match 39.4%; Score 13; DB 10; Length 6;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ISR 7  
111  
Db 2 ISR 4

RESULT 3

P82072

PRELIMINARY; PRT; 5 AA.

ID P82072;  
AC P82072;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
OS RUBELLIDIN 3.1.  
OS Litoria rubella (Desert tree frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litoria.  
OX NCBI\_TaxID=104895;  
RN [1]  
RP SEQUENCE, AND MASS SPECTROMETRY.  
RC TISSUE=SKIN SECRETION;  
RA Steinboerner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,  
RT Tyler M.J., Wallace J.C.;  
RT "The structure of new peptides from the Australian red tree frog  
Litoria rubella: the skin peptide profile as a probe for the study  
of evolutionary trends of amphibians".  
RL Aust. J. Chem. 49:955-963(1996).  
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR  
ANTIAPOTIC ACTIVITY.

CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
CC -1- MASS SPECTROMETRY: MW=655; METHOD=FMH.  
KW Amphibian skin; Amidation.  
FT MOD\_RES 5  
SQ SEQUENCE 5 AA: 656 MW: 71A9C9CB10300000 CRC64;

Query Match 33.3%; Score 11; DB 13; Length 5;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FT 2  
11  
Db 4 FT 5

RESULT 4

P82073

PRELIMINARY; PRT; 5 AA.

ID P82073;  
AC P82073;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
OS RUBELLIDIN 3.2.  
OS Litoria rubella (Desert tree frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litoria.  
OX NCBI\_TaxID=104895;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=SKIN SECRETION;  
RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;  
RT "Peptides from the skin glands of the Australian buzzing tree frog  
Litoria rubella. Comparison with the skin peptides from Litoria  
rubella".  
RL Aust. J. Chem. 52:0-0(1999).  
CC -1- FUNCTION: CAERIDINS SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR  
ANTIAPOTIC ACTIVITY.  
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.  
KW Amphibian skin.  
SQ SEQUENCE 5 AA: 570 MW: 71A9C9CB2A00000 CRC64;

Query Match 33.3%; Score 11; DB 13; Length 5;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FT 2  
11  
Db 4 FT 5

RESULT 5

O95945

PRELIMINARY; PRT; 7 AA.

ID O95945;  
AC O95945;  
DT 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
OS Sacccharomyces cerevisiae (Baker's yeast).  
OC Mitochondrion.  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D273-10B;  
RX MEDLINE=8106985; PubMed=6254986;  
RA Bonitz S.G., Cornuzzi G., Thalerfeld B., Tzagoloff A., Macino G.;  
RT "Assembly of the mitochondrial membrane system: Structure and  
nucleotide sequence of the gene coding for subunit 1 of yeast



```

RT cytochrome oxidase."
RL J. Biol. Chem. 255:11927-11941(1980).
DR EMBL: Y00694; CAA24066.1; -.
KW Mitochondrion.
FT NON_TER
SO SEQUENCE 1 1
      7 AA; 859 MW; 75B7232362CDC460 CRC64;

Query Match      33.3%; Score 11; DB 8; Length 7;
Best Local Similarity 66.7%; Pred. No. 5.6e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 KIS 6
   1 1
Db 5 KLS 7

RESULT 6
ID P82541 PRELIMINARY; PRT; 6 AA.
AC P82541;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CHLOROPLAST 30S RIBOSOMAL PROTEIN S19 BETA (FRAGMENT).
OS Spinacia oleracea (Spinach).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_TaxId=3562;
RN [1]
RP SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.
RC STRAIN=CV, AIMARO; TISSUE=LEAF;
RX MEDLINE=20435797; PubMed=10874039;
RA Yamaguchi K., von Knobilach K., Subramanian A.R.;
RT The plastid ribosomal proteins. Identification of all the proteins in
RT the small subunit of an organelle ribosome (chloroplast).";
RL J. Biol. Chem. 37:28455-28465(2000).
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES.
CC -1- MASS SPECTROMETRY: MW=10477.0; METHOD=ELECTROSPRAY.
CC -1- MASS SPECTROMETRY: MW=10495; METHOD=MALDI.
CC -1- MISCELLANEOUS: S19 ALPHA AND BETA FORMS DIFFER IN PI. S19 BETA
CC FORM IS THE MINOR BASIC FORM.
CC -1- MISCELLANEOUS: ON THE 2D-GEL TTS MW IS 12 KDA.
CC -1- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.
DR InterPro: IPR002222; Ribosomal_S19.
DR Pfam: PF00203; Ribosomal_S19; PARTIAL.
DR PRINTS: PR00975; RIBOSOMAL_S19; PARTIAL.
DR PROSITE: PS00323; RIBOSOMAL_S19; PARTIAL.
KW Ribosomal protein; Chloroplast; rRNA-binding.
FT NON_TER
SO SEQUENCE 6 AA; 732 MW; 633373735A411C000 CRC64;

Query Match      30.3%; Score 10; DB 10; Length 6;
Best Local Similarity 66.7%; Pred. No. 5.6e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 TLK 4
   1 1
Db 3 SLK 5

RESULT 7
ID O08433 PRELIMINARY; PRT; 4 AA.
AC O08433;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)

```

```

DE UDP-GLUCURONOSYLTRANSFERASE, MICROSOMAL (EC 2.4.1.17) (UDPGT)
DE (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10115;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GUNN;
RX MEDLINE=91282758; PubMed=1840486;
RA Sato H., Aono S., Kashimata S., Koike O.;
RT Genetic defect of bilirubin UDP-glucuronosyltransferase in the
RT hyperbilirubinemic Gunn rat.";
RL Biochem. Biophys. Res. Commun. 177:1161-1164(1991).
CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
CC ENDOGENOUS COMPOUNDS.
CC -1- CATALYTIC ACTIVITY: UDP-GLUCURONATE + ACCEPTOR = UDP + ACCEPTOR
CC BETA-D-GLUCURONOSIDE.
CC -1- SUBCELLULAR LOCATION: MICROSOME.
DR EMBL: S38636; AB019259.1;
KW transferase; glycosyltransferase; Microsome; Multigene family.
FT NON_TER
FT NON_TER
FT NON_TER
SO SEQUENCE 4 AA; 473 MW; 633732C420000000 CRC64;

Query Match      27.3%; Score 9; DB 11; Length 4;
Best Local Similarity 100.0%; Pred. No. 5.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 LK 4
   1 1
Db 3 LK 4

RESULT 8
ID O07354 PRELIMINARY; PRT; 7 AA.
AC O07354;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NIFK (FRAGMENT).
GN NIFK.
OC Synechococcus sp. (strain PCC 8801 / RF-1) (Cyanothecae PCC 8801).
OC Bacteria; Cyanobacteria; Chroococcales; Cyanolacce.
OX NCBI_TaxId=41431;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RF-1;
RX MEDLINE=99231861; PubMed=10217509;
RA Huang T.C., Lin R.F., Chu M.K., Chen H.M.;
RT "Organization and expression of nitrogen-fixation genes in the aerobic
RT nitrogen-fixing unicellular cyanobacterium Synechococcus sp. strain
RT RF-1.";
RL Microbiology 145:743-753(1999).
DR EMBL: AF003700; AAC35193.1; -.
FT NON_TER
FT NON_TER
SO SEQUENCE 7 AA; 849 MW; 7412C72AA9D5B030 CRC64;

Query Match      27.3%; Score 9; DB 2; Length 7;
Best Local Similarity 66.7%; Pred. No. 5.6e+05;
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTL 3
   1 1
Db 3 FDL 5

RESULT 9
ID O15857

```

ID 015897 PRELIMINARY: PRT: 7 AA.  
 AC 015897;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE (CLONE XP6A11a) (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PLACENTA;  
 RA Lee C.-C., Yazdani A., Wehnert M., Bailey J., Couch L., Xiong M.,  
 RA Coobanagh M.I., Chihault C.A., Baldini A., Lindsay E.A., Zhao Z.-Y.,  
 RA Caskey C.T.H.;  
 RT "Isolation of chromosome-specific genes by reciprocal probing of  
 RT arrayed cDNAs and cosmid libraries."  
 RL Hum. Mol. Genet. 0:0-0(1995).  
 DR EMBL: L32077; AAA73887.1; -.  
 FT NON\_TER 1 1  
 FT NON\_TER 7 7  
 SO SEQUENCE 7 AA: 814 MW: 672BID372046B0 CRC64;

Query Match 27.3%; Score 9; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 2: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 LK 4  
 DB 3 LK 4

RESULT 10  
 P93233  
 ID P93233 PRELIMINARY: PRT: 7 AA.  
 AC P93233;  
 DT 01-MAY-1997 (TREMBlrel. 03, Created)  
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE 1-AMINOACYLOPROPANE-1-CARBOXYLATE SYNTHASE (EC 4.4.1.14)  
 DE (FRAGMENT).  
 GN LE-ACSIB.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 NCBI\_Taxid=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=97351561; PubMed=9207843;  
 RA Oeliker J.H., Olson D.C., Shiu O.Y., Yang S.F.;  
 RT "Differential induction of seven 1-aminocyclopropane-1-carboxylate  
 RT synthase genes by elicitor in suspension cultures of tomato  
 RT (Lycopersicon esculentum)."  
 RL Plant Mol. Biol. 34:275-286(1997).  
 DR EMBL: U75692; AAC49682.1; -.  
 KW Lyase.  
 FT NON\_TER 1 1  
 FT NON\_TER 7 AA: 828 MW: 71B412C7377415D0 CRC64;

Query Match 27.3%; Score 9; DB 10; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 2: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 SR 7  
 DB 1 SR 2

RESULT 11

063480  
 ID 063480 PRELIMINARY: PRT: 7 AA.  
 AC 063480;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE TR4-NS ORPHAN RECEPTOR (FRAGMENT).  
 GN TR4.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=96198747; PubMed=8612486;  
 RA Yoshikawa T., Makino S., Geo X.M., Xing G.Q., Chuang D.M.,  
 RA Detera-Wadleigh S.D.;  
 RT "Splice variants of rat TR4 orphan receptor: differential expression  
 RT of novel sequences in the 5'-untranslated region and C-terminal  
 RT domain."  
 RL Endocrinology 137:1562-1571(1996).  
 DR EMBL: U59125; AAB02827.1; -.  
 FT NON\_TER 1 1  
 FT NON\_TER 7 AA: 758 MW: 672AAB87864005350 CRC64;

Query Match 27.3%; Score 9; DB 11; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 2: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KI 5  
 DB 1 KI 2

RESULT 12  
 055184  
 ID 055184 PRELIMINARY: PRT: 7 AA.  
 AC 055184;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE ORPHAN RECEPTOR TR4-NS (FRAGMENT).  
 GN TR4.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY;  
 RX MEDLINE=96198747; PubMed=8612486;  
 RA Yoshikawa T., Makino S., Geo X.M., Xing G.Q., Chuang D.M.,  
 RA Detera-Wadleigh S.D.;  
 RT "Splice variants of rat TR4 orphan receptor: differential expression  
 RT of novel sequences in the 5'-untranslated region and C-terminal  
 RT domain."  
 RL Endocrinology 137:1562-1571(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY;  
 RX MEDLINE=96299786; PubMed=8661150;  
 RA Yoshikawa T., Dupont B.R., Leach R.J., Detera-Wadleigh S.D.;  
 RT "New variants of the human and rat nuclear hormone receptor, TR4:  
 RT expression and chromosomal localization of the human gene."  
 RL Genomics 35:361-366(1996).  
 DR EMBL: U59454; AAB91433.1; -.  
 KW Receptor.  
 FT NON\_TER 1 1  
 FT NON\_TER 7 AA: 663 MW: 6DDAA8787EB05350 CRC64;

Query Match 27.3%; Score 9; DB 11; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KI 5  
 Db 1 KI 2

RESULT 13  
 ID 047029 PRELIMINARY; PRT; 7 AA.  
 AC 047029;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE AAD A1 PROTEIN (FRAGMENT).  
 GN AAD A1.  
 OS Enterobacter cloacae.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Enterobacter.  
 OX NCBI\_TaxID=550;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94079349; PubMed=8257126;  
 RA Rather P.N., Mann P.A., Mierzwa R., Hare R.S., Miller G.H., Shaw K.J.;  
 RT "Analysis of the aac(3)-Via gene encoding a novel 3-N-  
 RT acetyltransferase."  
 RL Antimicrob. Agents Chemother. 37:2074-2079(1993).  
 DR EMBL; M88012; AAA16193.1; -.  
 FT NON\_TER 1  
 SQ SEQUENCE 7 AA; 744 MW; 633862D2C321A030 CRC64;

Query Match 24.2%; Score 8; DB 2; Length 7;  
 Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
 Matches 1; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 KI 5  
 Db 3 KV 4

RESULT 14  
 ID P70804 PRELIMINARY; PRT; 7 AA.  
 AC P70804;  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ALGT PROTEIN (FRAGMENT).  
 GN ALGT.  
 OS Azotobacter vinelandii.  
 CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Azotobacter.  
 OX NCBI\_TaxID=354;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=E;  
 RX MEDLINE=96427318; PubMed=8830682;  
 RA Behm B.H.A., Ertesvag H., Valia S.;  
 RT "A new Azotobacter vinelandii mannuronan C-5-epimerase gene (alg6) is  
 RT part of an alg gene cluster physically organized in a manner similar  
 RT to that in Pseudomonas aeruginosa."  
 RL J. Bacteriol. 178:5884-5889(1996).  
 DR EMBL; X87973; CAA61230.1; -.  
 FT NON\_TER 1  
 SQ SEQUENCE 7 AA; 684 MW; 71B5A5A2D1AED0 CRC64;

Query Match 24.2%; Score 8; DB 2; Length 7;  
 Best Local Similarity 40.0%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 TLKIS 6  
 Db 2 TVSSS 6

RESULT 15  
 ID 09YIRO PRELIMINARY; PRT; 7 AA.  
 AC 09YIRO;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
 DE PVI CORE PROTEIN (FRAGMENT).  
 GN PVI.  
 OS Human adenovirus type 7a.  
 OS Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OX NCBI\_TaxID=85755;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KN 796-0620, S-1058, AND CL 68578;  
 RA Crawford-Miksza L.K., Nang R.N., Schnurr D.P.;  
 RT "Molecular surveillance of strain variation in adenoviruses causing  
 RT acute respiratory disease, AV 4 and AV 7a."  
 RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF065066; AAD03664.1; -.  
 DR EMBL; AF065066; AAD03664.1; -.  
 FT NON\_TER 1  
 SQ SEQUENCE 7 AA; 980 MW; 7B5EA414140322A0 CRC64;

Query Match 24.2%; Score 8; DB 12; Length 7;  
 Best Local Similarity 33.3%; Pred. No. 5.6e+05;  
 Matches 1; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 5 ISR 7  
 Db 1 VKR 3

Search completed: July 15, 2002, 13:37:39  
 Job time: 729 sec

11



```

Query Match      69.78;  Score 23;  DB 20;  Length 7;
Best Local Similarity 83.38;  PIdent No. 6.4e+05;
Matches 5;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;
OY      1 PTKIKIS 6
      2 fclsis 7
      3 fclsis 7
      4 fclsis 7
      5 fclsis 7
      6 fclsis 7
      7 fclsis 7
      8 fclsis 7
      9 fclsis 7
      10 fclsis 7
      11 fclsis 7
      12 fclsis 7
      13 fclsis 7
      14 fclsis 7
      15 fclsis 7
      16 fclsis 7
      17 fclsis 7
      18 fclsis 7
      19 fclsis 7
      20 fclsis 7
      21 fclsis 7
      22 fclsis 7
      23 fclsis 7
      24 fclsis 7
      25 fclsis 7
      26 fclsis 7
      27 fclsis 7
      28 fclsis 7
      29 fclsis 7
      30 fclsis 7
      31 fclsis 7
      32 fclsis 7
      33 fclsis 7
      34 fclsis 7
      35 fclsis 7
      36 fclsis 7
      37 fclsis 7
      38 fclsis 7
      39 fclsis 7
      40 fclsis 7
      41 fclsis 7
      42 fclsis 7
      43 fclsis 7
      44 fclsis 7
      45 fclsis 7
      46 fclsis 7
      47 fclsis 7
      48 fclsis 7
      49 fclsis 7
      50 fclsis 7
      51 fclsis 7
      52 fclsis 7
      53 fclsis 7
      54 fclsis 7
      55 fclsis 7
      56 fclsis 7
      57 fclsis 7
      58 fclsis 7
      59 fclsis 7
      60 fclsis 7
      61 fclsis 7
      62 fclsis 7
      63 fclsis 7
      64 fclsis 7
      65 fclsis 7
      66 fclsis 7
      67 fclsis 7
      68 fclsis 7
      69 fclsis 7
      70 fclsis 7
      71 fclsis 7
      72 fclsis 7
      73 fclsis 7
      74 fclsis 7
      75 fclsis 7
      76 fclsis 7
      77 fclsis 7
      78 fclsis 7
      79 fclsis 7
      80 fclsis 7
      81 fclsis 7
      82 fclsis 7
      83 fclsis 7
      84 fclsis 7
      85 fclsis 7
      86 fclsis 7
      87 fclsis 7
      88 fclsis 7
      89 fclsis 7
      90 fclsis 7
      91 fclsis 7
      92 fclsis 7
      93 fclsis 7
      94 fclsis 7
      95 fclsis 7
      96 fclsis 7
      97 fclsis 7
      98 fclsis 7
      99 fclsis 7
      100 fclsis 7

```

XX The present invention is concerned with producing scaffold proteins  
CC based upon the human CTLA-4 SCA domain. These scaffold proteins can be  
CC used as a scaffold to bind antigen- or receptor-binding fragments. These  
CC can be used in the treatment of diseases such as cancer,  
CC attherosclerosis, thrombosis, osteoporosis, rheumatoid arthritis and  
CC diabetic retinopathy. Sequences AAB29930-B29939 were used in the  
CC production of the proteins of the invention.  
XX

SQ Sequence 7 AA:

Query Match 69.7%; Score 23; DB 21; Length 7;  
Best Local Similarity 83.3%; Pred. No. 6,4e+05;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTKLTS 6  
|||  
Dd 2 fllsls 7

RESULT 3  
AAV40738  
ID AAV40738 standard; peptide; 7 AA.  
XX  
AC AAV40738:  
XX  
DT 01-DEC-1999 (first entry)  
XX  
DE S4 derivative #12, beta strand of scaffold protein structure.  
XX  
KW Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;  
KM tumour; chemotherapeutic agent.  
XX  
OS Synthetic.  
XX  
PN EP947582-A1.  
PM  
PD 06-OCT-1999.  
XX  
PF 31-MAR-1998; 98EP-0870065.  
PR 31-MAR-1998; 98EP-0870065.  
XX  
PA (INNO-) INNOGENETICS NV.  
XX  
PI Desmet J, Hufcon S, Hoogenboom H, Sablon E;  
XX WPI: 1999-542958/46.  
XX  
XX New scaffold protein, useful for stabilizing antigens used as vaccines  
PT -  
PS  
PS Disclosure; Page 6; 105pp; English.

Sequence AAV40727-Y40748 are functionally equivalent derivatives of the  
S4 peptide (AAI40607) which forms part of a scaffold protein. S4 is a  
beta strand peptide which forms part of a beta sheet. Peptides  
(AAV40601-Y40609) together form a single-chain scaffold protein which  
contains at least 1 disulfide bond, contains less than 10% alpha helix  
and contains at least 6 beta-strands. The scaffold protein is constructed  
of beta strands SI-S6, and may also include beta strands AI-A3, or any  
functionally equivalent derivative of these sequences. The beta strands  
form two beta sheets SI/S4/S3 and S6/S5/S2 with each strand connected to  
the next by hydrogen bonds, which generate a beta sandwich architecture.  
If the additional beta strands AI-A3 are included in the structure the  
scaffold is constructed of two beta sheets, with the structures  
AI/SI/S4/S3 and S6/S5/S2/AI/A3. The beta strands are connected to each  
receptor or antigen. The scaffold protein is used to stabilize antigens  
or whole proteins such as receptors, or their fragments. It may be used  
to bind two separate molecules. For example, one surface of the scaffold  
may be bound to a protein which binds to a tumour antigen. This will

CC target the complex to tumour cells. Another surface may be bound to a  
 CC cytotoxic molecule or an autoimmune antibody which may then kill the  
 CC tumour cells. Therefore the scaffold protein may be used to target  
 CC chemotherapeutic agents to specific cells. It may also be used to  
 CC stabilize individual peptides in a peptide library and may be used in  
 CC diagnostic techniques, and to stabilize antigens used as vaccines.

SO Sequence 7 AA;

Query Match 66.7%; Score 22; DB 20; Length 7;  
 Best Local Similarity 83.3%; Pred. No. 6.4e+05;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTUKIS 6  
 11111  
 Db 2 flltis 7

#### RESULT 4

AAB30076  
 ID AAB30076 standard; Peptide: 7 AA.

AC AAB30076;

DT 09-FEB-2001 (first entry)

DE Scaffold protein SCA S4 peptide SEQ ID NO: 137.

KW Human; CYLA-4; scaffold protein; antigen-binding; receptor-binding;  
 KW SCA domain; cancer; thrombosis; osteoporosis; rheumatoid arthritis;  
 KW diabetic retinopathy; atherosclerosis.

OS Synthetic.

PN WO200060070-A1.

PD 12-OCT-2000.

PF 01-APR-1999; 99WO-EP02283.

PR 01-APR-1999; 99WO-EP02283.

PA (INNO-) INNOGENETICS NV.

PI Desmet J, Hufon S, Hoogenboom H, Sablon E;

DR WPI; 2000-665002/64.

PT Scaffold composed of single-chain polypeptide having beta sandwich  
 PT architecture carrying new and randomized peptide sequences useful as  
 PT supporting framework and carrying antigen- or receptor binding  
 PT fragments -

PS Disclosure; Page 15; 68pp; English.

CC The present invention is concerned with producing scaffold proteins  
 CC based upon the human CYLA-4 SCA domain. These scaffold proteins can be  
 CC used as a scaffold to bind antigen- or receptor-binding fragments. These  
 CC can be used in the treatment of diseases such as cancer,  
 CC atherosclerosis, thrombosis, osteoporosis, rheumatoid arthritis and  
 CC diabetic retinopathy. Sequences AAB29930-A29939 were used in the  
 CC production of the proteins of the invention.

SO Sequence 7 AA;

Query Match 66.7%; Score 22; DB 21; Length 7;  
 Best Local Similarity 83.3%; Pred. No. 6.4e+05;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FTUKIS 6  
 11111

Db 2 flltis 7

#### RESULT 5

AAV42013  
 ID AAV42013 standard; Peptide: 7 AA.

AC AAV42013;

DT 09-DEC-1999 (first entry)

DE Rheumatoid arthritis diagnostic protein isoform peptide #164.

KW Rheumatoid arthritis diagnostic feature; RPI; synovial fluid;  
 KW rheumatoid arthritis diagnostic feature; RPI; synovial fluid;  
 KW rheumatoid arthritis diagnostic protein isoform; screening;  
 KW expression reference protein isoform; prognosis.

OS Homo sapiens.

PN WO9947925-A2.

PD 23-SEP-1999.

PF 15-MAR-1999; 99WO-GB00763.

PR 13-MAR-1998; 98GB-0005477.

PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.

PI Parekh RB, Patel TP, Townsend RR;

DR WPI; 1999-571871/48.

PT Diagnosis of human rheumatoid arthritis by two-dimensional  
 PT electrophoresis -

PS Disclosure; Page 21; 157pp; English.

CC A method has been developed for the diagnosis of human rheumatoid  
 CC arthritis (RA) using two-dimensional electrophoresis to generate a  
 CC two-dimensional array of features. The method can be used for screening,  
 CC diagnosis and prognosis of RA in a subject or for monitoring the effect  
 CC of an anti-RA drug or therapy administered to a subject. The method  
 CC comprises: (a) analysing a sample of serum or plasma and optionally  
 CC synovial fluid by two-dimensional electrophoresis, to generate a two-  
 CC dimensional array of features; (b) identifying at least one chosen  
 CC feature whose relative abundance correlates with the presence or absence  
 CC of RA; and (c) comparing the abundance of each chosen feature in the  
 CC sample with the abundance of that chosen feature in serum or plasma from  
 CC one or more persons without RA, where the relative abundance of the  
 CC chosen feature or features in the sample indicates the presence or  
 CC absence of RA in the subject. The method can also be used in clinical  
 CC studies for testing drugs for therapy of RA, for purification of RA-  
 CC diagnostic protein isoforms (RPIs), and for production of antibodies to  
 CC RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify  
 CC compounds that promote or inhibit their activity, which are then used as  
 CC RA drugs. Nucleic acid encoding RADFs can be used in gene therapy  
 CC protocols. AA141944 to AAV42100 represent RPI peptides, AAV42101 to  
 CC AAV42103 represent expression reference protein isoform peptides and  
 CC AA225066 to AA225068 represent degenerate probes for RPIs, which are all  
 CC used in the exemplification of the present invention.

SO Sequence 7 AA;

Query Match 63.6%; Score 21; DB 20; Length 7;  
 Best Local Similarity 83.3%; Pred. No. 6.4e+05;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 TUKISR 7  
 11111  
 Db 2 tlmstr 7

```

RESULT 6
AA81848
ID AA81848 standard; peptide: 7 AA.
XX
XX
AC AA81848;
XX
DT 16-MAY-1996 (first entry)
XX
DE Human afamin tryptic fragment FX20.
XX
XX
KW Human; afamin; serum protein family; albumin; alpha-fetoprotein; plasma;
KW Vitamin D binding protein; homology; post-translational processing;
KW Chromatography; Primer; PCR; amplification; probe; rheumatoid arthritis;
KW Ischaemia-reperfusion injury; ARDS; cardiopulmonary bypass; sepsis;
KW toxic plasma substance; inflammation.
XX
XX
OS Homo sapiens.
XX
PN MO9527059-A1.
XX
PD 12-OCT-1995.
XX
PF 31-MAR-1995; 95MO-US04075.
XX
PR 31-MAR-1994; 94US-0222619.
XX
PA (AMGE-) AMGEN INC.
PA (UTRO ) UNIV ROCKEFELLER.
XX
PI Lichenstein HS, Lyons DE, Wright SD, Wurfel MM;
XX
DR WPI: 1995-358634/46.
XX
XX
PT Human afamin or a variant and poly:nucleotide(s) encoding it - a
PT human serum protein with activities in common with other members of
PT this family.
XX
XX
PS Example 3; Page 45; 97pp; English.
XX
CC Peptides AA81847-54 are tryptic peptide fragments from human afamin
CC (AA81845) novel member of the human serum protein family. The
CC fragments were used to design primers and probes (AA800786-98) for the
CC cloning of the afamin gene (AA800785) from human liver cDNA. Afamin is
CC thought to have similar properties to human albumin, alpha-fetoprotein
CC and vitamin D binding protein due to homology with these proteins. The
CC gene encodes a mature protein of 66576 daltons without post-translational
CC processing (ca. 87000 daltons with post-translational processing). The
CC protein was isolated from human plasma by a conventional chromatographic
CC methods. The protein can be used to ameliorate ischaemia-reperfusion
CC injury, rheumatoid arthritis, ARDS, cardiopulmonary bypass, sepsis, toxic
CC plasma substances released after inflammation, etc.
XX
XX
SQ Sequence 7 AA:

```

```

Query Match 60.6%; Score 20; DB 16; Length 7;
Best Local Similarity 57.1%; Pred. No. 6.4e+05;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 FTLKISR 7
   1 1 1 1 1
DB 1 ftfeyer 7

```

```

RESULT 7
AAV41889
ID AAV41889 standard; peptide: 7 AA.
XX
XX
AC AAV41889;
XX
XX
DT 09-DEC-1999 (first entry)

```

```

XX
XX
DE Rheumatoid arthritis diagnostic protein isoform peptide #40.
XX
XX
KW Human; rheumatoid arthritis; RA; diagnosis; RPI; RAPD; detection;
KW rheumatoid arthritis diagnostic feature; ERPI; synovial fluid;
KW rheumatoid arthritis diagnostic protein isoform; screening;
KW expression reference protein isoform; prognosis.
XX
XX
OS Homo sapiens.
XX
PN WO9947925-A2.
XX
PD 23-SEP-1999.
XX
PF 15-MAR-1999; 99MO-GB00763.
XX
PR 13-MAR-1998; 98GB-0005477.
XX
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI Parekh RB, Patel TP, Townsend RR.
XX
DR WPI: 1999-571871/48.
XX
XX
PT Diagnosis of human rheumatoid arthritis by two-dimensional
PT electrophoresis -
XX
PS Disclosure: Page 18; 157pp; English.
XX
XX
CC A method has been developed for the diagnosis of human rheumatoid
CC arthritis (RA) using two-dimensional electrophoresis to generate a
CC two-dimensional array of features. The method can be used for screening,
CC diagnosis and prognosis of RA in a subject or for monitoring the effect
CC of an anti-RA drug or therapy administered to a subject. The method
CC comprises: (a) analysing a sample of serum or plasma and optionally
CC synovial fluid by two-dimensional electrophoresis, to generate a two-
CC dimensional array of features; (b) identifying at least one chosen
CC feature whose relative abundance correlates with the presence or absence
CC of RA; and (c) comparing the abundance of each chosen feature in the
CC sample with the abundance of that chosen feature in serum or plasma from
CC one or more persons without RA, where the relative abundance of the
CC chosen feature or features in the sample indicates the presence or
CC absence of RA in the subject. The method can also be used in clinical
CC studies for testing drugs for therapy of RA, for purification of RA-
CC diagnostic protein isoforms (RPIs), and for production of antibodies to
CC RPIs. The RA-diagnostic feature (RAPD) proteins can be used to identify
CC compounds that promote or inhibit their activity, which are then used as
CC RA drugs. Nucleic acid encoding RAPDs can be used in gene therapy
CC protocols. AAV41844 to AAV42100 represent RPI peptides. AAV42101 to
CC AAV42566 to AA22568 represent degenerate probes for RPIs, which are all
CC used in the exemplification of the present invention.
XX
XX
SQ Sequence 7 AA:

```

```

Query Match 60.6%; Score 20; DB 20; Length 7;
Best Local Similarity 42.9%; Pred. No. 6.4e+05;
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 FTLKISR 7
   1 1 1 1 1
DB 1 ytfelst 7

```

```

RESULT 8
ABB55870
ID ABB55870 standard; peptide: 7 AA.
XX
XX
AC ABB55870;
XX
XX
DT 15-FEB-2002 (first entry)

```



DE Vascular dementia-associated protein isoform (VPI) 70.  
XX  
KW Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;  
KW diagnosis; prognosis; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN MO200169261-A2.  
XX  
PD 20-SEP-2001.  
XX  
PF 14-MAR-2001; 2001WO-GB01106.  
XX  
PR 15-MAR-2000; 2000GB-0006285.  
PR 24-NOV-2000; 2000GB-0028734.  
PR 28-NOV-2000; 2000US-0724391.  
XX  
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX  
PI Herath HMC, Parekh RB, Rohlf C;  
XX  
DR WPI; 2001-557937/62.  
XX  
PT Screening, diagnosis or prognosis of vascular dementia (VD), useful for  
PT determining stage of VD and monitoring the effect of VD therapy, for  
PT comprises analysing body fluid by 2-dimensional electrophoresis for  
PT features correlated with VD -  
XX  
PS Claim 6; Page 31; 15pp; English.  
XX  
CC The invention relates to screening, diagnosis or prognosis of Vascular  
CC Dementia (VD) in a subject comprising analysing body fluid from the  
CC subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of  
CC features containing at least one chosen feature whose relative abundance  
CC correlates with the presence, absence, stage or severity of VD or  
CC predicts the onset or course of VD, especially detecting in a sample of  
CC cerebrospinal fluid (CSF) from the subject one of 223 VD-associated  
CC protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the  
CC specification. Detecting VD-associated features and VPI is useful for the  
CC screening, diagnosis or prognosis of VD, for determining the stage or  
CC severity of VD, for identifying a subject at risk of VD or for  
CC monitoring the effect of therapy administered to a subject having VD.  
CC Nucleic acids encoding a VPI or inhibiting the function of a VPI are  
CC useful for the treatment of VD and for gene therapy.  
XX  
SQ Sequence 7 AA;  
  
Query Match 60.6%; Score 20; DB 22; Length 7;  
Best Local Similarity 42.9%; Pred. No. 6,4e+05;  
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
OY 1 FTKISR 7  
Db 1 ytfelst 7  
  
RESULT 9  
ABB55981  
ID ABB55981 standard; Peptide; 7 AA.  
XX  
AC ABB55981;  
XX  
DT 15-FEB-2002 (first entry)  
XX  
DE Vascular dementia-associated protein isoform (VPI) 181.  
XX  
KW Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;  
KW diagnosis; prognosis; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN MO200169261-A2.  
FN

XX  
PD 20-SEP-2001.  
XX  
PF 14-MAR-2001; 2001WO-GB01106.  
XX  
PR 15-MAR-2000; 2000GB-0006285.  
PR 24-NOV-2000; 2000GB-0028734.  
PR 28-NOV-2000; 2000US-0724391.  
XX  
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX  
PI Herath HMC, Parekh RB, Rohlf C;  
XX  
DR WPI; 2001-557937/62.  
XX  
PT Screening, diagnosis or prognosis of vascular dementia (VD), useful for  
PT determining stage of VD and monitoring the effect of VD therapy, for  
PT comprises analysing body fluid by 2-dimensional electrophoresis for  
PT features correlated with VD -  
XX  
PS Claim 6; Page 33; 15pp; English.  
XX  
CC The invention relates to screening, diagnosis or prognosis of Vascular  
CC Dementia (VD) in a subject comprising analysing body fluid from the  
CC subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of  
CC features containing at least one chosen feature whose relative abundance  
CC correlates with the presence, absence, stage or severity of VD or  
CC predicts the onset or course of VD, especially detecting in a sample of  
CC cerebrospinal fluid (CSF) from the subject one of 223 VD-associated  
CC protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the  
CC specification. Detecting VD-associated features and VPI is useful for the  
CC screening, diagnosis or prognosis of VD, for determining the stage or  
CC severity of VD, for identifying a subject at risk of VD or for  
CC monitoring the effect of therapy administered to a subject having VD.  
CC Nucleic acids encoding a VPI or inhibiting the function of a VPI are  
CC useful for the treatment of VD and for gene therapy.  
XX  
SQ Sequence 7 AA;  
  
Query Match 60.6%; Score 20; DB 22; Length 7;  
Best Local Similarity 42.9%; Pred. No. 6,4e+05;  
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
OY 1 FTKISR 7  
Db 1 ytfelst 7  
  
RESULT 10  
ABB56283  
ID ABB56283 standard; Peptide; 7 AA.  
XX  
AC ABB56283;  
XX  
DT 15-FEB-2002 (first entry)  
XX  
DE Vascular dementia-associated protein isoform (VPI) 483.  
XX  
KW Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;  
KW diagnosis; prognosis; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN MO200169261-A2.  
XX  
PD 20-SEP-2001.  
XX  
PF 14-MAR-2001; 2001WO-GB01106.  
XX  
PR 15-MAR-2000; 2000GB-0006285.  
PR 24-NOV-2000; 2000GB-0028734.  
PR 28-NOV-2000; 2000US-0724391.  
PR

XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX Herath HMAC, Parekh RB, Rohlf C;  
PI WPI: 2001-557937/62.  
DR  
XX  
XX Screening, diagnosis or prognosis of vascular dementia (VD), useful for  
PT determining stage of VD and monitoring the effect of VD therapy,  
PT comprises analysing body fluid by 2-dimensional electrophoresis for  
PT features correlated with VD -  
XX  
XX Claim 6; Page 40; 151pp; English.  
XX  
XX The invention relates to screening, diagnosis or prognosis of Vascular  
CC Dementia (VD) in a subject comprising analysing body fluid from the  
CC subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of  
CC features containing at least one chosen feature whose relative abundance  
CC correlates with the presence, absence, stage or severity of VD or  
CC predicts the onset or course of VD, especially detecting in a sample of  
CC cerebrospinal fluid (CSF) from the subject one of 223 VD-associated  
CC protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the  
CC specification. Detecting VD-associated features and VPI is useful for the  
CC screening, diagnosis or prognosis of VD, for determining the stage or  
CC severity of VD, for identifying a subject at risk of VD or for  
CC monitoring the effect of therapy administered to a subject having VD.  
CC Nucleic acids encoding a VPI or inhibiting the function of a VPI are  
CC useful for the treatment of VD and for gene therapy.  
XX  
XX Sequence 7 AA:  
SO  
Query Match 60.6%; Score 20; DB 22; Length 7;  
Best Local Similarity 42.9%; Pred. No. 6.4e+05;  
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 1 FTLKISR 7  
| : : : : |  
Db 1 ytfelst 7  
RESULT 11  
ABB52190  
ID ABB52190 standard; Peptide: 7 AA.  
XX  
AC ABB52190;  
XX  
XX 08-FEB-2002 (first entry)  
DT  
XX  
XX Human API-146 tryptic digest peptide #1.  
DE  
XX  
XX Human; neuroprotective; nootropic; gene therapy; vaccine;  
KW Alzheimer's disease; Alzheimer's Disease-Associated Feature; AF;  
KW Alzheimer's Disease-Associated Protein Isoform; API; tryptic digest;  
KM Expression Reference Protein Isoform; ERPI; proteolysis.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200175454-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 03-APR-2001; 2001WO-US10908.  
PF  
XX  
XX 03-APR-2000; 2000US-194504P.  
PR  
XX  
XX 28-NOV-2000; 2000US-253647P.  
PR  
XX  
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
PA (PFI2 ) PRIZER INC.  
PA  
XX  
XX Durham KL, Friedman DL, Herath HMAC, Kimmel LH, Parekh RB;  
PI Potter DM, Rohlf C, Silber BM, Stiger TR, Sunderland PT;  
PI Townsend RR, White F, Williams SA;

XX WPI: 2001-639384/73.  
DR  
XX  
XX Screening for Alzheimer's disease in a mammal, by making  
PT two-dimensional array of a feature whose relative abundance correlates  
PT with disease, and comparing with abundance of the feature in samples of  
PT healthy persons -  
XX  
XX Example; Page 30; 162pp; English.  
XX  
XX The invention relates to methods for the screening, diagnosis and  
CC prognosis of Alzheimer's disease. The methods involve the detection  
CC of Alzheimer's Disease-Associated Features (AFs) and Alzheimer's  
CC Disease-Associated Protein Isoforms (APIs) in cerebrospinal fluid,  
CC serum or plasma. The abundance of the AFs and APIs is then  
CC normalised to an Expression Reference Protein Isoform (ERPI) in  
CC order to determine whether a patient is suffering from, or has  
CC a predisposition to, Alzheimer's Disease. The relative abundance of  
CC the AFs and APIs correlates with the severity of Alzheimer's Disease.  
CC The present sequence is a peptide produced from an API by proteolysis.  
XX  
XX Sequence 7 AA:  
SO  
Query Match 60.6%; Score 20; DB 22; Length 7;  
Best Local Similarity 57.1%; Pred. No. 6.4e+05;  
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 FTLKISR 7  
| : : : : |  
Db 1 ftfeyar 7  
RESULT 12  
ABB52355  
ID ABB52355 standard; Peptide: 7 AA.  
XX  
XX ABB52355;  
AC  
XX  
XX 08-FEB-2002 (first entry)  
DT  
XX  
XX Human API-125 tryptic digest peptide #8.  
DE  
XX  
XX Human; neuroprotective; nootropic; gene therapy; vaccine;  
KW Alzheimer's disease; Alzheimer's Disease-Associated Feature; AF;  
KW Alzheimer's Disease-Associated Protein Isoform; API; tryptic digest;  
KM Expression Reference Protein Isoform; ERPI; proteolysis.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200175454-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 03-APR-2001; 2001WO-US10908.  
PF  
XX  
XX 03-APR-2000; 2000US-194504P.  
PR  
XX  
XX 28-NOV-2000; 2000US-253647P.  
PR  
XX  
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
PA (PFI2 ) PRIZER INC.  
PA  
XX  
XX Durham KL, Friedman DL, Herath HMAC, Kimmel LH, Parekh RB;  
PI Potter DM, Rohlf C, Silber BM, Stiger TR, Sunderland PT;  
PI Townsend RR, White F, Williams SA;  
XX  
XX WPI: 2001-639384/73.  
DR  
XX  
XX Screening for Alzheimer's disease in a mammal, by making  
PT two-dimensional array of a feature whose relative abundance correlates  
PT with disease, and comparing with abundance of the feature in samples of  
PT healthy persons -  
XX

PS Example; Page 34; 162pp; English.  
XX  
CC The invention relates to methods for the screening, diagnosis and  
CC prognosis of Alzheimer's disease. The methods involve the detection  
CC of Alzheimer's Disease-Associated Features (AFs) and Alzheimer's  
CC Disease-Associated Protein Isoforms (APIs) in cerebrospinal fluid,  
CC serum or plasma. The abundance of the AFs and APIs is then  
CC normalised to an Expression Reference Protein Isoform (ERPI) in  
CC order to determine whether a patient is suffering from, or has  
CC a predisposition to, Alzheimer's Disease. The relative abundance of  
CC the AFs and APIs correlates with the severity of Alzheimer's Disease.  
CC The present sequence is a peptide produced from an API by proteolysis.  
XX  
SQ Sequence 7 AA;  
Query Match 60.6%; Score 20; DB 22; Length 7;  
Best Local Similarity 42.9%; Pred. No. 6.4e+05; Indels 0; Gaps 0;  
Matches 3; Conservative 3; Mismatches 1;  
QY 1 FTLKISR 7  
Db 1 ytfelst 7  
RESULT 13  
AAU28602 standard; Peptide; 7 AA.  
ID AAU28602;  
AC AAU28602;  
XX 03-JAN-2002 (first entry)  
DT  
DE DPI tryptic digest peptide #199.  
XX  
XX Human; depression associated protein isoform; tryptic digest peptide;  
KM DPI; cerebrospinal fluid; CSF; BAD; bipolar affective disorder;  
KM neuropsychiatric disorder; bipolar mood disorder; neuroleptic;  
KM maniac-depressive illness; schizoaffective disorder.  
XX  
XX Homo sapiens.  
OS  
XX WO200162787-A1.  
XX  
XX 30-AUG-2001.  
XX  
XX 23-FEB-2001; 2001WO-GB00786.  
XX  
XX 24-FEB-2000; 2000GB-0004412.  
XX 08-DEC-2000; 2000GB-0030050.  
XX 12-DEC-2000; 2000US-0254830.  
XX  
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX  
XX Herath HMAc, Parekh RB, Rohlf C, Terrett JA, Tyson KL;  
PI WPI; 2001-570626/64.  
XX  
XX Novel nucleic acid encoding a protein associated with bipolar affective  
XX disorder, which is used for diagnosis, prophylaxis and therapy of  
XX neuropsychiatric disorders, such as bipolar affective disorder -  
XX  
XX Disclosure; Page 34; 153pp; English.  
XX  
XX The present invention relates to the identification of depression  
XX associated protein isoforms (DPIs), particularly the tryptic digest  
XX peptides of these proteins. Some of the DPIs (AAU28404-AAU28625)  
XX described are decreased in the cerebrospinal fluid (CSF) of BAD  
XX (bipolar affective disorder) subjects, whilst other DPIs  
XX (AAU28626-AAU28887) are increased in BAD subjects. Also described  
XX are peptide sequences identified from DPI-45 and DPI-213 and the  
XX nucleic acid sequences they are encoded by. The sequences of these  
XX invention are useful for clinical screening, diagnosis, prognosis.

CC therapy and prophylaxis of neuropsychiatric disorders e.g. BAD (also  
CC known as bipolar mood disorder, BP), maniac-depressive illnesses,  
CC attention deficit disorders, schizoaffective disorders, and bipolar  
CC affective disorders. The present sequence represents one of the DPI  
CC tryptic digest peptides of the present invention.  
XX  
SQ Sequence 7 AA;  
Query Match 60.6%; Score 20; DB 22; Length 7;  
Best Local Similarity 42.9%; Pred. No. 6.4e+05; Indels 0; Gaps 0;  
Matches 3; Conservative 3; Mismatches 1;  
QY 1 FTLKISR 7  
Db 1 ytfelst 7  
RESULT 14  
AAU24969 standard; Peptide; 7 AA.  
ID AAU24969  
AC AAU24969;  
XX 18-DEC-2001 (first entry)  
DT  
DE Schizophrenia-Associated Protein Isoform (SPI) peptide #198.  
XX  
XX Schizophrenia-associated protein isoform; SPI; SPI-206; SPI-238; SPI-240;  
KM neuroleptic; gene therapy; cerebrospinal fluid; serum; plasma.  
XX  
XX Homo sapiens.  
OS  
XX WO200162785-A2.  
XX  
XX 30-AUG-2001.  
XX  
XX 23-FEB-2001; 2001WO-GB00792.  
XX  
XX 24-FEB-2000; 2000GB-0004415.  
XX 28-NOV-2000; 2000US-0750395.  
XX  
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX  
XX Herath HMAc, Parekh RB, Rohlf C, Terrett JA, Tyson KL;  
PI WPI; 2001-570624/64.  
XX  
XX New schizophrenia associated protein isoforms and encoding nucleic acid  
XX molecules, useful for treatment, diagnosis and prognosis of  
XX schizophrenia and screening for potential drugs for treatment and new  
XX drug targets -  
XX  
XX Disclosure; Page 32; 148pp; English.  
XX  
XX The sequence represents a schizophrenia-associated protein isoform (SPI).  
XX These protein isoforms, e.g. SPI-206, SPI-238 and SPI-240 are detectable  
XX in cerebrospinal fluid, serum or plasma and are useful markers of  
XX schizophrenia. The sequences can be used for treatment and diagnosis of  
XX schizophrenia, screening, prognosis, monitoring the results of therapy,  
XX identifying patients most likely to respond to a particular therapy and  
XX identification of new targets for drug treatment. SPI DNA is useful as a  
XX nucleic acid probe to detect the presence of nucleic acids or SPIs.  
XX  
XX Sequence 7 AA;  
Query Match 60.6%; Score 20; DB 22; Length 7;  
Best Local Similarity 42.9%; Pred. No. 6.4e+05; Indels 0; Gaps 0;  
Matches 3; Conservative 3; Mismatches 1;  
QY 1 FTLKISR 7  
Db 1 ytfelst 7

Db 1 ytfelst 7

RESULT 15

AAU26249 standard; Peptide; 7 AA.

AAU26249;

18-DEC-2001 (first entry)

Depression-Associated Protein Isoform DPI-208.

Human; Bipolar Affective Disorder; BAD; Depression-Associated feature;  
 DF; Depression-Associated protein isoform; DPI; Cerebro-spinal fluid;  
 CSF; antidepressant; antimanic; nootropic; tranquilliser; neuroleptic;  
 attention deficient disorder; schizoaffective disorder;  
 unipolar affective disorder.

Homo sapiens.

MO200163294-A2.

30-AUG-2001.

23-FEB-2001; 2001MO-GB00791.

24-FEB-2000; 2000GB-0004412.

08-DEC-2000; 2000GB-0030050.

12-DEC-2000; 2000US-0234830.

(OXPO-) OXFORD GLYCOSCIENCES UK LTD.

Herath HMAC, Parekh RB, Rohlf C;

WPI; 2001-582081/65.

Preparation for diagnosing or treating bipolar affected disorder (BAD)  
 or unipolar depression, or for screening for modulators, comprises a  
 BAD-associated protein isoform -  
 Claim 8; Page 34; 163pp; English.

The invention relates to a preparation comprising an isolated Bipolar  
 affected disorder (BAD)-Associated Protein Isoform (DPIs). The DPIs are  
 used to screen, diagnose or prognose of BAD or unipolar depression,  
 determine the stage or severity of BAD or unipolar depression, identify a  
 subject at risk of developing BAD or unipolar depression, or monitor the  
 effect of therapy in a subject. They are also used to screen for or  
 identify agents that interact with a DPI. These agents, antibodies  
 against the DPIs, and nucleic acids encoding the DPIs are used to treat  
 or prevent BAD or unipolar depression. Diseases that can be treated are  
 attention deficient disorder, a schizoaffective disorder, a bipolar or a  
 unipolar affective disorder. The DPIs are used in proteomics. The  
 proteomic approach of using DPIs for screening, diagnosis or prognosis of  
 BAD or unipolar depression overcomes the problems of using gene  
 expression analysis, such as not being able to obtain central nervous  
 system (CNS) tissue from a living patient under normal circumstances.  
 The present sequence is a DIP decreased in the CSF (cerebro-spinal  
 fluid) of subjects having BAD.

Sequence 7 AA;

Query Match

Best Local Similarity 60.6%; Score 20; DB 22; Length 7;  
 Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY \*1 FTLKISR 7

Db 1 ytfelst 7

Search completed: July 15, 2002, 13:25:27  
 Job time: 1458 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:26:14 ; Search time 40.89 seconds  
(without alignments)  
4.181 Million cell updates/sec

Title: US-09-712-819A-6  
Perfect score: 33  
Sequence: 1 FTLKISR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 39160

Minimum DB seq length: 0  
Maximum DB seq length: 7

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA: \*  
1: /cgn2\_6/prodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/prodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/prodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/prodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/prodata/2/1aa/PCFUS.COMB.pep: \*  
6: /cgn2\_6/prodata/2/1aa/Backfilest.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	60.6	7	1 US-08-222-619-8	Sequence 8, Appl
2	20	60.6	7	5 PCT-US95-04075-8	Sequence 8, Appl
3	19	57.6	6	5252328-8	Patent No. 5252328
4	18	54.5	7	1 US-08-346-333-24	Sequence 24, Appl
5	18	54.5	7	5 US-09-268-992-78	Sequence 78, Appl
6	18	54.5	7	5 PCT-US91-07506-24	Sequence 24, Appl
7	17	51.5	7	2 US-09-174-060-17	Sequence 17, Appl
8	17	51.5	7	2 US-08-654-623-69	Sequence 69, Appl
9	17	51.5	7	3 US-08-338-382-17	Sequence 17, Appl
10	16	48.5	5	4 US-08-591-632-23	Sequence 23, Appl
11	16	48.5	6	1 US-07-717-331F-5	Sequence 5, Appl
12	16	48.5	7	1 US-07-634-641-12	Sequence 12, Appl
13	16	48.5	7	1 US-08-136-743B-55	Sequence 55, Appl
14	16	48.5	7	3 US-09-040-216-28	Sequence 28, Appl
15	16	48.5	7	4 US-09-173-941-72	Sequence 72, Appl
16	15	45.5	4	1 US-08-305-871A-17	Sequence 17, Appl
17	15	45.5	5	1 US-08-136-743B-63	Sequence 63, Appl
18	15	45.5	5	3 US-09-040-216-55	Sequence 55, Appl
19	15	45.5	5	4 US-08-591-632-17	Sequence 17, Appl
20	15	45.5	5	4 US-08-591-632-26	Sequence 26, Appl
21	15	45.5	6	1 US-08-136-743B-62	Sequence 62, Appl
22	15	45.5	6	1 US-08-357-264-6	Sequence 6, Appl
23	15	45.5	6	1 US-08-297-731-4	Sequence 4, Appl
24	15	45.5	6	1 US-08-222-619-21	Sequence 21, Appl
25	15	45.5	6	1 US-08-290-448A-41	Sequence 41, Appl
26	15	45.5	6	1 US-08-672-514-6	Sequence 6, Appl
27	15	45.5	6	1 US-08-290-448A-41	Sequence 41, Appl

28	15	45.5	6	1 US-08-175-069A-41	Sequence 41, Appl
29	15	45.5	6	2 US-08-637-759B-219	Sequence 219, App
30	15	45.5	6	3 US-08-871-355A-219	Sequence 219, App
31	15	45.5	6	3 US-09-040-216-54	Sequence 54, Appl
32	15	45.5	6	3 US-09-623-618B-6	Sequence 6, Appl
33	15	45.5	6	4 US-08-461-939B-41	Sequence 41, Appl
34	15	45.5	6	4 US-08-464-000-41	Sequence 41, Appl
35	15	45.5	6	4 US-09-201-945-219	Sequence 219, App
36	15	45.5	6	4 US-09-187-859-650	Sequence 650, App
37	15	45.5	6	4 US-09-187-859-2845	Sequence 2845, App
38	15	45.5	6	4 US-09-187-859-2908	Sequence 2908, App
39	15	45.5	6	4 US-09-187-859-3264	Sequence 3264, App
40	15	45.5	6	4 US-09-187-859-3174	Sequence 3174, App
41	15	45.5	6	5 PCT-US95-04075-21	Sequence 21, Appl
42	15	45.5	6	5 PCT-US95-10793-4	Sequence 4, Appl
43	15	45.5	6	5 5342922-11	Patent No. 5342922
44	15	45.5	7	1 US-08-136-743B-6	Sequence 6, Appl
45	15	45.5	7	1 US-08-136-743B-26	Sequence 26, Appl

## ALIGNMENTS

```
RESULT 1
US-08-222-619-8
; Sequence 8, Application US/08222619
; Patent No. 5652352
;
GENERAL INFORMATION:
; APPLICANT: Lichenstein, Henri
; APPLICANT: Lyons, David
; APPLICANT: Wurfel, Mark
; APPLICANT: Wright, Samuel
; TITLE OF INVENTION: Afatinib: A Human Serum Albumin-Like
; NUMBER OF INVENTION: Protein
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Amgen Center, Patent Operations/RRC
; STREET: 1840 Dehavenland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: U.S.
; ZIP: 91320-1789
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,619
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-222-619-8

Query Match      60.6%; Score 20; DB 1; Length 7;
Best Local Similarity 57.1%; Pred. No. 1.7e+05;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 FTLKISR 7
       11: 11
Db      1 FTLEYSR 7

RESULT 2
PCT-US95-04075-8
; Sequence 8, Application PC/US9504075
```

```

; GENERAL INFORMATION:
; APPLICANT: AMGEN INC.
; TITLE OF INVENTION: Alamin: A Human Serum Albumin-Like
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Amgen Center, Patent Operations/RCR
; STREET: 1840 DeHavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: U.S.
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04075
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ. ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; PCT-US95-04075-8

Query Match          60.6%; Score 20; DB 5; Length 7;
Best Local Similarity 57.1%; Pred. No. 1.7e+05;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FTLKISR 7
   ||: ||
   1 FTFEYSR 7
   ||: ||

RESULT 3
5252328-8
; Patent No. 5252328
; APPLICANT: FAULDS, DARYL, VISHOOT, MIMI, BROOKS, EMILY
; TITLE OF INVENTION: MYCOPLASMA HOPDNEUMONIAE ANTIGEN AND USES
; THEREFORE
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/335,726
; FILING DATE: 07-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 30,130
; FILING DATE: 26-MAR-1987
; SEQ ID NO: 8:
; LENGTH: 6
5252328-8

Query Match          57.6%; Score 19; DB 6; Length 6;
Best Local Similarity 80.0%; Pred. No. 1.7e+05;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FTLKI 5
   |||
   2 FVLKI 6
   |||

RESULT 4
US-08-346-333-24
; Sequence 24, Application US/08346333
; Patent No. 5677153
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
```

```

; APPLICANT: Palzkill, Timothy
; TITLE OF INVENTION: Methods for modifying DNA and for
; TITLE OF INVENTION: detecting effects of such modification on interaction of
; TITLE OF INVENTION: encoded modified polypeptides with target substrates.
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Richard F. Treccartin
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/346,333
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/039,501
; FILING DATE:
; APPLICATION NUMBER: US/07/602,158
; FILING DATE: 22-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ. ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-346-333-24

Query Match          54.5%; Score 18; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TLKI 5
   |||
   2 TLKI 5
   |||

RESULT 5
US-09-268-992-78
; Sequence 78, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-10p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/086,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
```

SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 78  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-268-992-78

Query Match 54.5%; Score 18; DB 4; Length 7;  
Best Local Similarity 50.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FTLKIS 6  
: : : : :  
Db 2 YTMKYS 7

RESULT 6  
PCT-US91-07506-24  
Sequence 24, Application PC/ITUS9107506  
GENERAL INFORMATION:  
APPLICANT: Borstein, David  
APPLICANT: Palzikill, Timothy  
TITLE OF INVENTION: Methods for modifying DNA and for  
TITLE OF INVENTION: detecting effects of such modification on interaction of  
TITLE OF INVENTION: encoded modified polypeptides with target substrates.  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Richard F. Trecartin  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/07506  
FILING DATE: 19911021  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Trecartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: FP-53469-PC/RF  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US91-07506-24

Query Match 54.5%; Score 18; DB 5; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTKI 5  
: : : : :  
Db 2 TTKI 5

RESULT 7  
US-09-174-060-17  
Sequence 17, Application US/09174060  
Patent No. 5989554

GENERAL INFORMATION:  
APPLICANT: Knuth, Mark W  
APPLICANT: Haak-Frendscho, Mary  
APPLICANT: Shultz, John W  
APPLICANT: Lesley, Scott A  
APPLICANT: Villars, Catherine E  
TITLE OF INVENTION: HIGH LEVEL EXPRESSION AND FACILE  
TITLE OF INVENTION: PURIFICATION OF PROTEINS, PEPTIDES AND CONJUGATES FOR  
TITLE OF INVENTION: IMMUNIZATION, PURIFICATION AND DETECTION APPLICATIONS  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ross & Stevens, S.C.  
STREET: 1 South Pinckney St.  
CITY: Madison  
STATE: WI  
COUNTRY: USA  
ZIP: 53701

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/174,060  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/338,382  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sara, Charles S  
REGISTRATION NUMBER: 30,492  
REFERENCE/DOCKET NUMBER: 34506.024  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-257-5353  
TELEFAX: 608-257-9175  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-174-060-17

Query Match 51.5%; Score 17; DB 2; Length 7;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLK 4  
: : : : :  
Db 1 YTLK 4

RESULT 8  
US-08-654-623-69  
Sequence 69, Application US/08654623  
Patent No. 6010884  
GENERAL INFORMATION:  
APPLICANT: Griffiths, Andrew D  
APPLICANT: Holliger, Kaspar-Philipp  
APPLICANT: Nissim, Ahuva  
APPLICANT: Fisch, Igor  
APPLICANT: Winter, Gregory P  
TITLE OF INVENTION: Recombinant Binding Proteins and Peptides  
NUMBER OF SEQUENCES: 71  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/654,623  
FILING DATE: 29-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: (C12N 1/21, C12R 1.19)  
FILING DATE: 04-DEC-1992  
APPLICATION NUMBER: GB 9225453.1  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9300816.7  
FILING DATE: 16-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 93303614.7  
FILING DATE: 10-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9319969.3  
FILING DATE: 22-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/02492  
FILING DATE: 03-DEC-1993  
APPLICATION NUMBER: GB 9412147.2  
FILING DATE: 17-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB94/02662  
FILING DATE: 05-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/448,418  
FILING DATE: 02-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: David W. Clough  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 28111/33259  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-654-623-69

Query Match 51.5%; Score 17; DB 3; Length 7;  
Best Local Similarity 60.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

2 TLKTS 6  
1 SLKVS 5

RESULT 9  
US-08-338-382-17  
Sequence 17, Application US/08338382  
Patent No. 6069230  
GENERAL INFORMATION:  
APPLICANT: Knuth, Mark W  
APPLICANT: Haak-Frendscho, Mary  
APPLICANT: Shultz, John W  
APPLICANT: Lesley, Scott A  
APPLICANT: Villars, Catherine E  
TITLE OF INVENTION: HIGH LEVEL EXPRESSION AND FACILE  
TITLE OF INVENTION: PURIFICATION OF PROTEINS, PEPTIDES AND CONJUGATES FOR  
TITLE OF INVENTION: IMMUNIZATION, PURIFICATION AND DETECTION APPLICATIONS  
NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ross & Stevens, S.C.  
STREET: 1 South Pinckney St.  
CITY: Madison  
STATE: WI  
COUNTRY: USA  
ZIP: 53701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/338,382  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sara, Charles S  
REGISTRATION NUMBER: 30,492  
REFERENCE/DOCKET NUMBER: 34506,024  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-257-5353  
TELEFAX: 608-257-9175  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-338-382-17

Query Match 51.5%; Score 17; DB 3; Length 7;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 ETLK 4  
1 TLK 4

1 ETLK 4  
1 TLK 4

RESULT 10  
US-08-591-632-23  
Sequence 23, Application US/08591632  
Patent No. 6261558  
GENERAL INFORMATION:  
APPLICANT: Bardas, Carlos F.  
APPLICANT: Burton, Dennis R.  
APPLICANT: Lerner, Richard A.  
TITLE OF INVENTION: SYNTHETIC HUMAN NEUTRALIZING MONOCLONAL  
TITLE OF INVENTION: ANTIBODIES TO HUMAN IMMUNODEFICIENCY VIRUS  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The Scripps Research Institute, Office of  
ADDRESSEE: Patent Counsel  
STREET: 10550 No. 6261558th Torrey Pines Road, TPC 8  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/591,632  
FILING DATE: 19-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/11907  
FILING DATE: 19-OCT-1994  
PRIOR APPLICATION DATA:



APPLICATION NUMBER: US 08/308,841  
FILING DATE: 19-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/233,619  
FILING DATE: 26-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/139,409  
FILING DATE: 19-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitting, Thomas  
REGISTRATION NUMBER: 34,163  
REFERENCE/DOCKET NUMBER: TSRI 332.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 784-2937  
TELEFAX: (619) 784-9399  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-591-632-23

Query Match 48.5%; Score 16; DB 4; Length 5;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLK 4  
111:  
DB 2 FTIQ 5

RESULT 11  
US-07-717-331F-5  
Sequence 5, Application US/07/717331F  
Patent No. 5484905  
GENERAL INFORMATION:  
APPLICANT: June Nasrallah; Michael Nasrallah; and Joshua  
APPLICANT: Stein  
TITLE OF INVENTION: A Receptor Protein Kinase Gene  
TITLE OF INVENTION: Encoded At The Self-Incompatibility Locus  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Yahwak & Associates  
STREET: 25 Skytop Drive  
CITY: Trumbull  
STATE: Connecticut  
COUNTRY: USA  
ZIP: 06611  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/717,331F  
FILING DATE: June 19th 1991  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: George M. Yahwak  
REGISTRATION NUMBER: 26,824  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (203)268-1951  
TELEFAX: (203)268-1951  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-717-331F-5

Query Match 48.5%; Score 16; DB 1; Length 6;  
Best Local Similarity 75.0%; Pred. No. 1.7e+05;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LKIS 6  
111:  
DB 2 LKVS 5

RESULT 12  
US-07-634-641-12  
Sequence 12, Application US/07634641  
Patent No. 5386011  
GENERAL INFORMATION:  
APPLICANT: Wiedeman, Paul E.  
APPLICANT: Kawai, Megumi  
APPLICANT: Iuliy, Jay R.  
APPLICANT: Or, Val-Sun  
APPLICANT: Magner, Rolf  
TITLE OF INVENTION: Hexa- and Heptapeptide Anaphylatoxin  
TITLE OF INVENTION: Receptor Ligands  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: One Abbott Park Road  
CITY: No. 5386011th Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/634,641  
FILING DATE: 19901227  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Janssen, Jerry F.  
REGISTRATION NUMBER: 29,175  
REFERENCE/DOCKET NUMBER: 4934.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 937-4558  
TELEFAX: (708) 937-9556  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-634-641-12

Query Match 48.5%; Score 16; DB 1; Length 7;  
Best Local Similarity 28.6%; Pred. No. 1.7e+05;  
Matches 2; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 FTLKIS 7  
111:  
DB 1 FRRILGR 7

RESULT 13  
US-08-136-743B-55  
Sequence 55, Application US/08136743B  
Patent No. 5459063  
GENERAL INFORMATION:  
APPLICANT: Barry S. Cooperman, Harvey Rubin,  
APPLICANT: Jerome Salem, and Allison L. Fisher  
TITLE OF INVENTION: "Plasmodium falciparum Ribonu-

```

; TITLE OF INVENTION: Gcleotide Reductase, DNA Sequences Therefor and Peptide Inhibit
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The University of Pennsylvania
; STREET: Suite 330
; STREET: 3700 Market Street
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19104-3246
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/136,743B
; FILING DATE: 10/14/93
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 3957-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEX: No. 5459063e
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
US-08-136-743B-55

Query Match      48.5%; Score 16; DB 1; Length 7;
Best Local Similarity 60.0%; Pred. NO. 1.7e+05;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FTLKI 5
   ||| :
Db 1 FTLDL 5

RESULT 14
US-09-040-216-28
; Sequence 28; Application US/09040216
; Patent No. 6030942
; GENERAL INFORMATION:
; APPLICANT: COOPERMAN, ET AL., BARRY
; TITLE OF INVENTION: PEPTIDES, PEPTIDE ANALOGS, PEPTIDOMIMETICS, AND OTHER
; TITLE OF INVENTION: SMALL MOLECULES USEFUL FOR INHIBITING THE ACTIVITY OF
; TITLE OF INVENTION: RIBONUCLEOTIDE REDUCTASE
; FILE REFERENCE: 9596-6301
; CURRENT APPLICATION NUMBER: US/09/040,216
; CURRENT FILING DATE: 1998-03-17
; EARLIER APPLICATION NUMBER: 08/919,748
; EARLIER FILING DATE: 1997-08-28
; EARLIER APPLICATION NUMBER: 60/025,146
; EARLIER FILING DATE: 1996-08-30
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: Ribonucleotide reductase inhibitor peptide
; FEATURE:
; OTHER INFORMATION: residue 1: MOD_RBS: ACETYLATION
;
US-09-040-216-28
```

```

Query Match      48.5%; Score 16; DB 3; Length 7;
Best Local Similarity 60.0%; Pred. NO. 1.7e+05;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FTLKI 5
   ||| :
Db 1 FTLDL 5

RESULT 15
US-09-173-941-72
; Sequence 72; Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV00815
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 72
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
; OTHER INFORMATION: codon binding sequence
;
US-09-173-941-72

Query Match      48.5%; Score 16; DB 4; Length 7;
Best Local Similarity 75.0%; Pred. NO. 1.7e+05;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 KISR 7
   ||| :
Db 4 KLSR 7
```

Search completed: July 15, 2002, 13:26:15  
Job time: 495 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 15, 2002, 13:41:35 ; Search time 14.1 seconds  
(without alignments)  
61.334 Million cell updates/sec

Title: US-09-712-819A-7

Perfect score: 43  
Sequence: 1 TDFLTRISS 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 788

Minimum DB seq length: 0  
Maximum DB seq length: 9

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR.71:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	22	51.2	8	T14906	hypothetical prote
2	16	37.2	8	A23967	leucopyrokinin - M
3	16	37.2	9	A28924	fructose-bisphosph
4	14	32.6	8	S71919	alcohol dehydrogen
5	14	32.6	9	A24244	adipokinetic hormo
6	14	32.6	9	G85802	hypothetical prote
7	13	30.2	6	P00028	pev-kinin 2 - pena
8	13	30.2	8	S11545	adipokinetic hormo
9	13	30.2	8	S13661	polygalacturonase
10	13	30.2	9	B38740	ig kappa chain C r
11	12	27.9	5	A32516	cholecystokinin-5
12	12	27.9	5	PT0644	T-cell receptor be
13	12	27.9	7	B39127	phosphotransferase
14	12	27.9	7	E30608	ig kappa chain V-I
15	12	27.9	7	PT0665	T-cell receptor be
16	12	27.9	8	S10596	adipokinetic hormo
17	12	27.9	8	S55310	adipokinetic hormo
18	12	27.9	8	A58620	adipokinetic hormo
19	12	27.9	8	P00012	cholecystokinin -
20	12	27.9	8	A43001	cholecystokinin -
21	12	27.9	8	PL0184	capsid protein VP-
22	12	27.9	8	J50318	leucokinin VITI -
23	12	27.9	8	H41978	caliIFMRamide 8 -
24	12	27.9	8	E47393	neuropeptide calla
25	12	27.9	8	S65381	cytochrome-c oxida
26	12	27.9	8	S66296	Na+-transporting A
27	12	27.9	9	A61357	phyllocaerulein -
28	12	27.9	9	A43848	cell surface adhes
29	12	27.9	9	S36898	ribosomal protein

30	12	27.9	9	2	A41978	caliIFMRamide 1 -
31	12	27.9	9	2	A44787	caliIFMRamide 10 -
32	12	27.9	9	2	B41978	caliIFMRamide 2 -
33	12	27.9	9	2	C41978	caliIFMRamide 3 -
34	12	27.9	9	2	D41978	caliIFMRamide 4 -
35	12	27.9	9	2	E41978	caliIFMRamide 5 -
36	12	27.9	9	2	F41978	caliIFMRamide 6 -
37	12	27.9	9	2	G41978	caliIFMRamide 7 -
38	12	27.9	9	2	A45199	L-tyrosophorin - Ja
39	12	27.9	9	2	PT0299	ig heavy chain CRD
40	12	25.6	4	2	I40697	biotin A - citroba
41	11	25.6	4	2	A40135	branched-chain-am
42	11	25.6	5	2	B44817	35k structural p
43	11	25.6	5	2	D44817	peptidyl-di
44	11	25.6	8	2	A42523	hypertrehalosemic
45	11	25.6	8	2	S08996	

#### ALIGNMENTS

RESULT 1  
T14906  
hypothetical protein - parsley  
C/Species: Petroselinum crispum (parsley)  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C/Accession: T14906  
R:Feldbrugge, M.; Sprenger, M.; Dinkelbach, M.; Yazaki, K.; Harter, K.; Weisshaar, B.  
Plant Cell 6, 1607-1621, 1994  
A>Title: Functional analysis of a light-responsive plant bZIP transcriptional regulat  
A/Reference number: T14906  
A/Accession: T14906  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-8 <FEL>  
A/Cross-references: EMBL:S75395; NID:g913201; PIR:e194245

Query Match 51.2% Score 22; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TLRIS 8  
DB 4 TLRIS 8

RESULT 2  
A23967  
leucopyrokinin - Madeira cockroach  
C/Species: Leucophaea madeira (Madeira cockroach)  
C/Date: 31-Mar-1988 #sequence\_revision 26-May-1994 #text\_change 11-Jul-1997  
C/Accession: A23967  
R:Nachman, R.J.; Holman, G.M.; Cook, B.J.  
Biochem. Biophys. Res. Commun. 137, 936-942, 1986  
A>Title: Active fragments and analogs of the insect neuropeptide leucopyrokinin: stru  
A/Reference number: A23967; MUID:86265041  
A/Accession: A23967  
A/Molecule type: protein  
A/Residues: 1-8 <MAC>  
C/Keywords: amidated carboxyl end; neuropeptide; pyroglutamic acid  
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
F:8/Modified site: amidated carboxyl end (Leu) #status experimental

Query Match 37.2% Score 16; DB 2; Length 8;  
Best Local Similarity 75.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TDFPT 4  
DB 2 TSFT 5

RESULT 3  
A28924  
fructose-bisphosphate aldolase (EC 4.1.2.13) B, hepatic - rabbit (fragment)  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 28-Aug-1989 #sequence\_revision 28-Aug-1989 #text\_change 07-Feb-1997  
C:Accession: A28924  
R:Jacko, A.G.; Brox, L.W.; Gracy, R.W.; Horecker, B.L.  
J. Biol. Chem. 245, 2140-2141, 1970  
A:Title: The carboxyl-terminal structure of rabbit liver aldolase (aldolase B).  
A:Reference number: A28924; MUID:70166720  
A:Accession: A28924  
A:Molecule type: protein  
A:Residues: 1-9 <IAK>  
C:Keywords: aldehyde-lyase; carbon-carbon lyase; gluconeogenesis; glycolysis; liver; per

Query Match  
Best Local Similarity 37.2%; Score 16; DB 2; Length 9;  
Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 FTLTSS 9  
1 1 1 1  
DB 2 FLATTSS 8

RESULT 4  
S71919  
alcohol dehydrogenase (EC 1.1.1.1) - grass carp (fragment)  
C:Species: Ctenopharyngodon idella (grass carp)  
C:Date: 14-Apr-1998 #sequence\_revision 24-Apr-1998 #text\_change 07-May-1999  
C:Accession: S71919  
R:Tsuji, H.T.; Mock, W.Y.; Lau, K.K.; Fong, W.P.  
Biochem. Biophys. Acta 1296, 41-46, 1996  
A:Title: Proteolytic activation of grass carp (Ctenopharyngodon idellus) liver alcohol de  
A:Reference number: S71919; MUID:96350418  
A:Accession: S71919  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-8 <TSU>  
A:Note: the source is designated Ctenopharyngodon idellus  
C:Keywords: NAD; oxidoreductase

Query Match  
Best Local Similarity 32.6%; Score 14; DB 2; Length 8;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TDFT 4  
1 1 1 1  
DB 5 TRFT 8

RESULT 5  
A24244  
adipokine hormone - bollworm  
N:Alternate names: Hez-AKH  
C:Species: Heliothis zea (bollworm, corn earworm, tomato fruitworm)  
C:Date: 31-Mar-1988 #sequence\_revision 23-Mar-1995 #text\_change 31-Oct-1997  
C:Accession: A24244  
R:Jaffe, H.; Rains, A.K.; Riley, C.T.; Fraser, B.A.; Holman, G.M.; Wagner, R.M.; Ridgway  
Biochem. Biophys. Res. Commun. 135, 622-628, 1986  
A:Title: Isolation and primary structure of a peptide from the corpora cardiaca of Helio  
A:Reference number: A24244; MUID:86186794  
A:Accession: A24244  
A:Molecule type: protein  
A:Residues: 1-9 <IAF>  
C:Superfamily: adipokine hormone  
C:Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic  
F:1/Modified site: pyroglutamate carboxylic acid (Gln) #status experimental  
F:9/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match  
Best Local Similarity 32.6%; Score 14; DB 2; Length 9;  
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 LTISS 9  
1 1 1 1  
DB 2 LTFTS 6

RESULT 6  
G85802  
hypothetical protein 22947 [imported] - Escherichia coli (strain 0157:H7, substrain E  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: G85802  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
iller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: G85802  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-9 <STO>  
A:Cross-references: GB:AE005174; NID:g12515957; PIDN:AAG36883.1; GSPDB:GN00145; UWGP:  
A:Experimental source: strain 0157:H7, substrain EDL933  
C:Genetics:  
A:Gene: 22947

Query Match  
Best Local Similarity 32.6%; Score 14; DB 2; Length 9;  
Matches 2; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 FTLTSS 9  
1 1 1 1  
DB 3 YTFMLSN 9

RESULT 7  
PD0028  
pev-kinin 2 - penaeid shrimp (Penaeus vannamei) (fragment)  
C:Species: Penaeus vannamei  
C:Date: 21-Aug-1998 #sequence\_revision 21-Aug-1998 #text\_change 19-May-2000  
C:Accession: PD0028  
R:Nieto, J.; Veelaert, D.; Denu, R.; Maekens, E.; Cerstiaens, A.; Coast, G.; Devree  
Biochem. Biophys. Res. Commun. 248, 406-411, 1998  
A:Title: Identification of one tachykinin- and two kinin-related peptides in the brai  
A:Reference number: PD0027; MUID:98342103  
A:Accession: PD0028  
A:Molecule type: protein  
A:Residues: 1-6 <NIB>  
C:Comment: This peptide belongs to myotropic neuropeptides.

Query Match  
Best Local Similarity 30.2%; Score 13; DB 2; Length 6;  
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 DFT 4  
1 1 1 1  
DB 1 DFS 3

RESULT 8  
S11545  
adipokine hormone - nestling-sucking blowfly  
C:Species: Protophormia terreovae (nestling-sucking blowfly)  
C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 31-Oct-1997  
C:Accession: S11545  
R:Gaede, G.; Wilps, H.; Kellner, R.  
Biochem. J. 269, 309-313, 1990  
A:Title: Isolation and structure of a novel charged member of the red-pigment-concent

eriaceae (Diptera).  
 A:Reference number: S11545; MUID:90351345  
 A:Accession: S11545  
 A:Molecule type: protein  
 A:Residues: 1-8 <GAE>  
 C:Superfamily: adipokine hormone  
 C:Keywords: amidated carboxyl end; corpora cardiaca; hormone; neuropeptide; pyroglutamic  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F:8/Modified site: amidated carboxyl end (Trp) #status experimental

Query Match 30.2%; Score 13; DB 2; Length 8;  
 Best Local Similarity 75.0%; Pred. No. 2.8e+05;  
 Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 LTTS 8  
 1 1 1  
 DB 2 LTFS 5

RESULT 9  
 S13661  
 polygalacturonase (EC 3.2.1.15) isoform PG2 - fungus (Sclerotinia sclerotiorum) (fragment)  
 N:Alternate names: endopolygalacturonase; pectin depolymerase; pectinase  
 C:Species: Sclerotinia sclerotiorum  
 C:Date: 19-Mar-1997 #sequence\_revision 21-Nov-1998 #text\_change 21-Nov-1998  
 C:Accession: S13661  
 R:Waksman, G.; Keon, J. P. R.; Turner, G.  
 Biochim. Biophys. Acta 1073, 43-48, 1991  
 A:Title: Purification and characterization of two endopolygalacturonases from Sclerotinia  
 A:Reference number: S13661; MUID:91120822  
 A:Accession: S13661  
 A:Molecule type: protein  
 A:Residues: 1-8 <WAK>  
 C:Function:  
 A:Description: Involved in pectin degradation  
 C:Keywords: glycosidase; hydrolase

Query Match 30.2%; Score 13; DB 2; Length 8;  
 Best Local Similarity 60.0%; Pred. No. 2.8e+05;  
 Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 LTTS 8  
 1 1 1  
 DB 3 TTFPS 7

RESULT 10  
 B38740  
 Ig kappa chain C region (PY20) - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 19-Mar-1997 #sequence\_revision 13-Mar-1998 #text\_change 13-Mar-1998  
 C:Accession: B38740  
 R:Ruff-Jamison, S.; Campos-Gonzalez, R.; Glenney Jr., J. R.  
 J. Biol. Chem. 266, 6607-6613, 1991  
 A:Title: Heavy and light chain variable region sequences and antibody properties of anti  
 A:Reference number: A38740; MUID:91177923  
 A:Accession: B38740  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-9 <RUF>

Query Match 30.2%; Score 13; DB 2; Length 9;  
 Best Local Similarity 37.5%; Pred. No. 2.8e+05;  
 Matches 3; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 DPTTIS 9  
 1 1 1  
 DB 2 DAAPTVGS 9

## RESULT 11

A32516  
 cholecystokinin-5 - dog  
 N:Alternate names: CCK-5  
 C:Species: Canis lupus familiaris (dog)  
 C:Date: 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 18-Aug-2000  
 C:Accession: A32516  
 R:Shively, J.; Reeve Jr., J. R.; Eysselein, V. E.; Ben-Avram, C.; Vigna, S. R.; Walsh, J.  
 Am. J. Physiol. 252, G272-G275, 1987  
 A:Title: CCK-5: sequence analysis of a small cholecystokinin from canine brain and in  
 A:Reference number: A32516; MUID:87153871  
 A:Accession: A32516  
 A:Molecule type: protein  
 A:Residues: 1-5 <SHR>  
 C:Comment: This peptide corresponds to the five carboxyl-terminal residues of cholecy  
 C:Superfamily: gastrin  
 C:Keywords: amidated carboxyl end; neuropeptide  
 F:5/Modified site: amidated carboxyl end (Phe) #status experimental

Query Match 27.9%; Score 12; DB 2; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DF 3  
 1 1  
 DB 4 DF 5

RESULT 12  
 PT0544  
 T-cell receptor beta chain V-D-J region (111-16) - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
 C:Accession: PT0544  
 R:Feeney, A. J.  
 J. Exp. Med. 174, 115-124, 1991  
 A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions  
 A:Reference number: PT0509; MUID:91277601  
 A:Accession: PT0544  
 A:Status: translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-5 <FEF>  
 A:Experimental source: newborn thymus, strain BALB/c  
 C:Keywords: T-cell receptor

Query Match 27.9%; Score 12; DB 2; Length 5;  
 Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TDTT 4  
 1 1 1  
 DB 2 SSTT 5

RESULT 13  
 B39127  
 phosphotransferase system enzyme II (EC 2.7.1.65) - Escherichia coli (fragment)  
 C:Species: Escherichia coli  
 C:Date: 27-Nov-1991 #sequence\_revision 27-Nov-1991 #text\_change 08-Oct-1999  
 C:Accession: B39127  
 R:Hardesty, C.; Ferran, C.; DiRienzo, J. M.  
 J. Bacteriol. 173, 449-456, 1991  
 A:Title: Plasmid-mediated sucrose metabolism in Escherichia coli: characterization of  
 A:Reference number: A39127; MUID:91100329  
 A:Accession: B39127  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-7 <HAR>  
 A:Cross-references: GB:M38416; NID:9155142; PIDN:AAA9418.1; PID:9155144  
 C:Keywords: phosphotransferase

Query Match 27.9%; Score 12; DB 2; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0;

QY 2 DF 3  
 ||  
 Db 2 DF 3

## RESULT 14

E30608  
 Ig kappa chain V-III region (Gag) - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 29-Jun-1989 #sequence\_revision 29-Jun-1989 #text\_change 16-Aug-1996  
 C:Accession: E30608  
 R:Goni, F.R.; Chen, P.P.; McGinnis, D.; Arjonilla, M.L.; Fernandez, J.; Carson, D.; Sold  
 J. Immunol. 142, 3158-3163, 1989  
 A:Title: Structural and idiotypic characterization of the L chains of human IgM autoantib  
 A:Reference number: A30601; MUID:89215279  
 A:Accession: E30608  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-7 <GON>  
 C:Keywords: heterotetramer; immunoglobulin

Query Match 27.9%; Score 12; DB 2; Length 7;  
 Best Local Similarity 42.9%; Pred. No. 2.8e+05;  
 Matches 3; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DF 3  
 ||  
 Db 1 EIVLTQS 7

## RESULT 15

PT0665  
 T-cell receptor beta chain V-D-J region (121-38M) - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
 C:Accession: PT0665  
 R:Feeney, A.J.  
 J. Exp. Med. 174, 115-124, 1991  
 A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
 A:Reference number: PT0509; MUID:91277601  
 A:Accession: PT0665  
 A:Status: translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-7 <FEF>  
 A:Experimental source: day 4 postnatal thymus, strain BALB/c  
 C:Keywords: T-cell receptor

Query Match 27.9%; Score 12; DB 2; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DF 3  
 ||  
 Db 6 DF 7

Search completed: July 15, 2002, 13:43:36  
 Job time: 121 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:43:21 ; Search time 10.19 Seconds  
(without alignments)  
34.198 Million cell updates/sec

Title: US-09-712-819A-7

Perfect score: 43

Sequence: 1 TDFYTLTSS 9

Scoring table: BLOSUM62

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 231

```

Maximum DB seq length: 9
Minimum DB seq length: 0

```

MAXIMUM OF 25% LENGTH.

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17	39.5	8	1	LMT2_LOCOMI	P22396 locusta milt
2	17	39.5	9	1	UPA3_HUMAN	P30089 homo sapien
3	16	37.2	8	1	LPR_LEIMA	P13049 leucophaea
4	15	34.9	9	1	UPH2_HUMAN	P40929 homo sapien
5	14	32.6	7	1	ALL7_CYPPO	P82156 cydia pomon
6	14	32.6	8	1	ALL5_CYPPO	P82158 cydia pomon
7	14	32.6	9	1	CAD1_ENFPA	P13268 entercoccc
8	14	32.6	8	1	DI_NEPNO	P24816 neptrops no
9	12	27.9	5	1	RE11_LITRU	P82070 litloria rub
10	12	27.9	7	1	GRP_MOUSE	P89025 mus muscul
11	12	27.9	8	1	ARK_LITBU	P23418 libellula c
12	12	27.9	8	1	ALL5_CALVO	P41841 calliphora
13	12	27.9	8	1	CCKN_MAEU	P30369 macropus el
14	12	27.9	8	1	COXG_RAT	P80430 rattus norv
15	12	27.9	8	1	FAR8_CALVO	P41863 calliphora
16	12	27.9	8	1	FUSS_FUSSO	P81010 fusarium so
17	12	27.9	8	1	LCK8_LEIMA	P13990 leucophaea
18	12	27.9	9	1	PARI_CALVO	P41856 calliphora
19	12	27.9	9	1	PAR2_CALVO	P41857 calliphora
20	12	27.9	9	1	PAR3_CALVO	P41858 calliphora
21	12	27.9	9	1	FAR4_CALVO	P41859 calliphora
22	12	27.9	9	1	FAR5_CALVO	P41860 calliphora
23	12	27.9	9	1	FAR6_CALVO	P41861 calliphora
24	12	27.9	9	1	FAR7_CALVO	P41862 calliphora
25	12	27.9	9	1	FARA_CALVO	P41865 calliphora
26	11	25.6	5	1	BIOA_CIFER	P13071 citrobacte
27	11	25.6	5	1	RE31_LITRU	P82072 litloria rub
28	11	25.6	5	1	RE32_LITRU	P82073 litloria rub
29	11	25.6	8	1	AKH_TIRBAT	P14595 tabanus alb
30	11	25.6	8	1	HFE2_PERAM	P04549 periplanet
31	11	25.6	8	1	ORNY_ORCIT	P82455 orotomeces
32	11	25.6	9	1	CCAP_CARNA	P38556 carclaus m
33	11	25.6	9	1	PGIR_DIAB	P81179 diaptreses

34	10	23.3	5	1	ET04_LITRBU	P82100_litoria_rub
35	10	23.3	8	1	LPMS_STAPR	P23211_staphylococ
36	10	23.3	9	1	LPCA_STRAU	P36884_staphylococ
37	9	20.9	7	1	CCF1_ENTFA	P20104_enterococcu
38	9	20.9	7	1	C1A_ENTFA	P11932_enterococcu
39	9	20.9	8	1	A1L1_CYPDO	P82152_cydia pomon
40	9	20.9	8	1	A1L2_CYPDO	P82157_cydia pomon
41	9	20.9	8	1	UPRA_HUMAN	P30096_homo sapien
42	9	20.9	9	1	BUR_CLOPA	P81337_clostridium
43	9	20.9	9	1	ISOT_CYPCA	P42933_cyprinus ca
44	9	20.9	1		MGMT_BOVIN	P29177_bos taurus
45	9	20.9	1		NSKI_SABRU	P41492_sarcophaga

## ALIGNMENTS

RESULT	1
ID	LMT2_LOCMI
STANDARD:	PRT: 8 AA.
AC	P2396;
DT	01-AUG-1991 (Rel. 19, Created)
DT	01-AUG-1991 (Rel. 19, Last sequence update)
DE	01-AUG-1991 (Rel. 19, Last annotation update)
LOCUS	Locustamytropin 2 (LOM-MT-2).
OC	Locusta migratoria (migratory Locust).
CC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
CC	Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
CC	Acridomorpha; Acridoidea; Acrididae; Locusta.
OX	NCBI_TaxID=7004;
RN	[1]
RP	SEQUENCE.
RC	TISSUE=Corpora cardiaca;
RA	Schoofs L., Holman G.M., Hayes T.K., Nachman R.J., de Loof A.;
RT	*Isolation, identification and synthesis of locustamytropin II, an
RT	additional neuropeptide of Locusta migratoria. Member of the
RL	cephalomyotropic peptide family.?
RL	Insect Biochem. 20:479-484(1990).
-I	FUNCTION: MEDIATES VISCERAL MUSCLE CONTRACTILE ACTIVITY
-I	(MYOTROPIC ACTIVITY).
-I	SIMILARITY: BELONGS TO THE PYROKININ FAMILY.
DR	Interpro: IPRO01484; PYROKININ.
DR	PROSITE: PS00539; PYROKININ_1.
KW	Neuropeptide; Amidation; Pyrokinin.
FT	MOD_RES 8 AMIDATION.
SQ	SEQUENCE 8 AA: 934 MW: 26341771A9CAA87B CRC64;
Query Match	39.5%; Score 17; DB 1; Length 8;
Best Local Similarity	100.0%; Pred. No. le=05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
OY	2 DFT 4
DB	3 DFT 5
RESULT 2	
ID	UPA3_HUMAN
STANDARD:	PRT: 9 AA.
AC	P30089;
DT	01-APR-1993 (Rel. 25, Created)
DT	01-APR-1993 (Rel. 25, Last sequence update)
DE	15-OCT-2001 (Rel. 40, Last annotation update)
DE	Unknown protein from 2D-page of plasma (Spot 11) (Fragment).
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE.
RC	TISSUE=Plasma;
KX	MEDLINE=93092937; PubMed=1459097;

RA Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquall C.,  
 RA Sanchez J.-C., James R., Tisot J.-D., Bjellqvist B.,  
 RA Hochstrasser D.F.,  
 RT Plasma protein map: an update by microsequencing."  
 RL Electrophoresis 13:707-714(1992).  
 CC -1- MISCELLANEOUS: ON THE 2D-GEL, THE DETERMINED PI OF THIS UNKNOWN  
 CC PROTEIN IS: 4.6, ITS MW IS: 46 KDa.  
 DR SWISS-2DPAGE: P30089; HUMAN.  
 FT NON\_TER 1  
 FT NON\_TER 9  
 SQ SEQUENCE 9 AA: 1056 MW; 26F2B1BA769C737 CRC64;

Query Match 39.5%; Score 17; DB 1; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDF 3  
 DB 7 TDF 9

RESULT 3  
 LPK\_LEUMA STANDARD; PRT; 8 AA.  
 ID LPK\_LEUMA  
 AC P13049;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Leukoprotektin (LPK) (LEW-PK).  
 OS Leucophaea maderae (Madelira cockroach).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;  
 OC Blaberoidae; Blaberidae; Leucophaea.  
 OX NCBI\_TaxID=6988;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=86269041; PubMed=3015140;  
 RA Nachman R.J., Holman G.M., Cook B.J.,  
 RT "Active fragments and analogs of the insect neuropeptide  
 RT leucoprotektin: structure-function studies."  
 RL Biochem. Biophys. Res. Commun. 137:936-942(1986).  
 RN [2]  
 RP SEQUENCE, AND SYNTHESIS.  
 RC TISSUE-Head;  
 RX MEDLINE=87052651; PubMed=2877794;  
 RA Holman G.M., Cook B.J., Nachman R.J.,  
 RT "Primary structure and synthesis of a blocked myotropic  
 RT neuropeptide isolated from the cockroach, Leucophaea maderae."  
 RL Comp. Biochem. Physiol. 85C:219-224(1986).  
 CC -1- FUNCTION: MEDIATES VISCERAL MUSCLE CONTRACTILE ACTIVITY  
 CC (MYOTROPIC ACTIVITY).  
 CC -1- MISCELLANEOUS: AN ANALOG WITHOUT THE N-TERMINAL POA RESIDUE WAS  
 CC SYNTHESIZED AND FOUND TO EXHIBIT GREATER ACTIVITY (144%) THAN THE  
 CC PARENT NEUROPEPTIDE. THE PORTION OF THE SEQUENCE OF LPK MOST  
 CC CRITICAL FOR THE MYOTROPIC PROPERTIES IS LIMITED TO THE  
 CC PENTAPEPTIDE FRAGMENT FTPL.  
 CC -1- SIMILARITY: BELONGS TO THE PYROKININ FAMILY.  
 DR InterPro: IPR001484; Pyrokinin.  
 DR PROSITE: PS00539; PYROKININ; 1.  
 KW Neuropeptide; Amidation; Pyrokinin.  
 FT MOD\_RES 1  
 FT MOD\_RES 8  
 FT MOD\_RES 1  
 FT MOD\_RES 8  
 SQ SEQUENCE 8 AA: 949 MW; 92341771A9D5A1B6 CRC64;

Query Match 37.2%; Score 16; DB 1; Length 8;  
 Best Local Similarity 75.0%; Pred. No. 1e+05;  
 Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TDF 4  
 DB 1 TDF 1

DB 2 TSFT 5

RESULT 4  
 UHA2\_HUMAN STANDARD; PRT; 9 AA.  
 ID UHA2\_HUMAN  
 AC P40929;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Unknown protein from 2D-page of heart (Spot 5603) (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE-Heart;  
 RX MEDLINE=95203287; PubMed=7895732;  
 RA Corbett J.M., Wheeler C.H., Baker C.S., Yacoub M.H., Dunn M.J.,  
 RT "The human myocardial two-dimensional gel protein database: update  
 RT 1994."  
 RL Electrophoresis 15:1459-1465(1994).  
 CC -1- MISCELLANEOUS: ON THE 2D-GEL, THE DETERMINED PI OF THIS UNKNOWN  
 CC PROTEIN IS: 6.0, ITS MW IS: 55.3 KDa.  
 FT NON\_TER 9  
 SQ SEQUENCE 9 AA: 1104 MW; 8874B1B5B01B2CA CRC64;

Query Match 34.9%; Score 15; DB 1; Length 9;  
 Best Local Similarity 50.0%; Pred. No. 1e+05;  
 Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDF 4  
 DB 4 TEXT 7

RESULT 5  
 ALL7\_CYDPO STANDARD; PRT; 7 AA.  
 ID ALL7\_CYDPO  
 AC P82158;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Cydastactin 7.  
 OS Cydia pomonella (Coddling moth).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Tortricidae; Tortricidae; Olethreutinae; Cydia.  
 OX NCBI\_TaxID=82600;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE-Larva;  
 RX MEDLINE=98054539; PubMed=9392829;  
 RA Davey H., Johnson A.H., Maesturo J.-L., Scott A.G., Winstanley D.,  
 RA Davey M., East P.D., Thorpe A.,  
 RT "Lepidopteran peptides of the allatostatin superfamily."  
 RL Peptides 18:1301-1309(1997).  
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 KW Neuropeptide; Amidation.  
 FT MOD\_RES 7  
 FT MOD\_RES 7  
 SQ SEQUENCE 7 AA: 873 MW; 672879CABB569350 CRC64;

Query Match 32.6%; Score 14; DB 1; Length 7;  
 Best Local Similarity 75.0%; Pred. No. 1e+05;  
 Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 DFTL 5  
 DB 4 DFTL 7



```

RESULT 6
ID ALL5_CYPDPO STANDARD: PRT: 8 AA.
AC P82156;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cydiastatin 5
OS Cydia pomonella (Codling moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Tortricoidae; Tortricidae; Olethreutinae; Cydia.
OX NCBI_TaxID=82600;
RN [1]
RP SEQUENCE.
RC TISSUE=Larva;
RX MEDLINE=98054539; PubMed=9392829;
RA Duve H., Johnson A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA Davey M., East P.D., Inorpe A.;
RT "Lepidopteran peptides of the allatostatin superfamily.";
RL Peptides 18:1301-1309(1997).
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW Neuropeptide; Amidation.
FT MOD.RES 8
SQ SEQUENCE 8 AA; 898 MW; 922879CAB58640D CRC64;

```

```

Query Match 32.6%; Score 14; DB 1; Length 8;
Best Local Similarity 75.0%; Pred. No. 1e+05;
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 2 DFTL 5
DB 5 DFGL 8

```

```

RESULT 7
ID CADI_ENTFA STANDARD: PRT: 8 AA.
AC P13268;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 01-FEB-1991 (Rel. 17, Last annotation update)
DE Sex pheromone CADI.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE.
RX MEDLINE=85051889; PubMed=6437872;
RA Moril M., Segakami Y., Narita M., Isogai A., Fujino M., Kitada C.,
RA Craig R.A., Clewell D.B., Suzuki A.;
RT "Isolation and structure of the bacterial sex pheromone, CADI, that
RT induces plasmid transfer in Streptococcus faecalis.";
RL FEMS Lett. 178:97-100(1984).
CC -1- FUNCTION: CADI IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
CC HEMOLYSIN PLASMID PADI.
KW Pheromone.
SQ SEQUENCE 8 AA; 819 MW; 047DD732C735B9C7 CRC64;

```

```

Query Match 32.6%; Score 14; DB 1; Length 8;
Best Local Similarity 33.3%; Pred. No. 1e+05;
Matches 2; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 3 FTLTIS 8
DB 2 FSLVLA 7

```

```

RESULT 8

```

```

ID1 NEPNO STANDARD: PRT: 9 AA.
AC P24816;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DE 01-MAR-1992 (Rel. 21, Last annotation update)
DE Gastrin/cholecystokinin-like peptide D1.
OS Nephrops norvegicus (Norway lobster).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Nephropidae; Nephropidae; Nephrops.
OX NCBI_TaxID=6829;
RN [1]
RP SEQUENCE.
RC TISSUE=Stomach;
RX MEDLINE=92082847; PubMed=1747388;
RA Favrel P., Kegel G., Sedlmeyer D., Keller R., van Wormhoudt A.;
RT "Structure and biological activity of crustacean gastrointestinal
RT peptides identified with antibodies to gastrin/cholecystokinin.";
RL Biochimie 73:1233-1239(1991).
CC -1- FUNCTION: MAY CONTROL DIGESTION PROCESSES IN CRUSTACEANS.
CC -1- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
DR PIR: A48398; A48398.
KW Hormone.
SQ SEQUENCE 9 AA; 1038 MW; 60EC79CAB6D8787B CRC64;

```

```

Query Match 32.6%; Score 14; DB 1; Length 9;
Best Local Similarity 75.0%; Pred. No. 1e+05;
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 2 DFTL 5
DB 6 DFVL 9

```

```

RESULT 9
ID REL1_LITRU STANDARD: PRT: 5 AA.
AC P82070;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Rubellidin 1.1.
OS Litorea rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Litorea.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE=Skin secretion;
RA Steinbomer S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,
RA Tyler M.J., Wallace J.C.;
RT "The structure of new peptides from the Australian red tree frog
RT Litorea rubella. The skin peptide profile as a probe for the study
RT of evolutionary trends of amphibians.";
RL Aust. J. Chem. 49:955-963(1996).
CC -1- FUNCTION: SHOW NEITHER NEUROPEPTIDE ACTIVITY NOR ANTIBIOTIC
CC ACTIVITY.
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN DORSAL GLANDS.
CC -1- MASS SPECTROMETRY: MM=598; METHOD=FAB.
KW Amphibian skin.
SQ SEQUENCE 5 AA; 598 MW; 6DD9C9CAB2A00000 CRC64;

```

```

Query Match 27.9%; Score 12; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 DF 3
DB 2 DF 3

```

RESULT 10  
 GFRP\_MOUSE STANDARD; PRT; 7 AA.  
 AC P99025;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE GTP cyclohydrolase I feedback regulatory protein (P35) (Fragment).  
 GN GCFRP OR GFRP  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Liver;  
 RA Sanchez J.-C., Rouse V., Frutiger S., Hughes G., Yan J.X.,  
 RA Hoogland C., Appel R.D., Blinz P.A., Hochstrasser D.F.,  
 RA Courtner M., Appel R.D., Blinz P.A., Hochstrasser D.F.,  
 RL Submitted (Aug-1998) to the SWISS-PROT data bank.  
 CC -1- FUNCTION: MEDIATES TETRAHYDROBIPTERIN INHIBITION OF GTP  
 CC CYCLOHYDROLASE 1. THIS INHIBITION IS REVERSED BY L-PHENYLMALANINE  
 CC (BY SIMILARITY).  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC SWISS-2DPAGE: P99025; MOUSE.  
 DR INT MET 0  
 FT NON TER 0  
 SO SEQUENCE 7 AA; 806 MW; 7185B05727384700 CRC64;

Query Match 27.9%; Score 12; DB 1; Length 7;  
 Best Local Similarity 60.0%; Pred. No. 1e+05;  
 Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 LTSS 9  
 DB 3 LLIST 7

RESULT 11  
 AKH\_LIBAU STANDARD; PRT; 8 AA.  
 ID AKH\_LIBAU  
 AC P25418;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Adipokineic hormone (AKH).  
 OS Libellula auripennis (Skimmer dragonfly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Palaeoptera; Odonata; Anisoptera; Libellulidae; Libellula.  
 OX NCBI\_TaxID=6966;  
 RN [1]  
 RP SEQUENCE, AND SYNTHESIS.  
 RC TISSUE=Corpora cardiaca;  
 RX MEDLINE=90359055; PubMed=2390213;  
 RA Gaede G.;  
 RT "The putative ancestral peptide of the adipokineic/red-pigment-  
 RT concentrating hormone family isolated and sequenced from a  
 RT dragonfly.";  
 RL Biol. Chem. Hoppe-Seyler 371:475-483(1990).  
 CC -1- FUNCTION: THIS HORMONE, RELEASED FROM CELLS IN THE CORPORA  
 CC CARDIACA AFTER THE BEGINNING OF FLIGHT, CAUSES RELEASE OF  
 CC DIGESTIBLES FROM THE FAT BODY AND THEN STIMULATES THE FLIGHT  
 CC MUSCLES TO USE THESE DIGESTIBLES AS AN ENERGY SOURCE.  
 CC -1- SIMILARITY: BELONGS TO THE AKH / HRTN / RPCH FAMILY.  
 DR InterPro: IPR002047; AKH.  
 DR PROSITE: PS00256; AKH; 1.  
 KM Neuropeptide; Amidation; Flight.  
 FT MOD\_RES 1 1  
 FT MOD\_RES 8 8 AMIDATION.

SO SEQUENCE 8 AA; 978 MW; 8665A771A9C452D6 CRC64;

Query Match 27.9%; Score 12; DB 1; Length 8;  
 Best Local Similarity 66.7%; Pred. No. 1e+05;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 DFT 4  
 DB 3 NPT 5

RESULT 12  
 ALLS\_CALVO STANDARD; PRT; 8 AA.  
 ID ALLS\_CALVO  
 AC P41841;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Callistostatin 5 (Met-callistostatin 1) ([Hyp3]Met-callistostatin).  
 OS Calliphora vomitoria (Blue blowfly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Oestridae; Calliphoridae; Calliphora.  
 OX NCBI\_TaxID=27454;  
 RN [1]  
 RP SEQUENCE.  
 RC MEDLINE=93211980; PubMed=8460157;  
 RA Dube H., Johnsen A.H., Scott A.G., Yu C.G., Yagi K.J., Tobe S.S.,  
 RA Thorpe A.;  
 RT "Callistostatin: neuropeptides from the blowfly Calliphora vomitoria  
 RT with sequence homology to cockroach allatostatin.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:2456-2460(1993).  
 RN [2]  
 RP CHARACTERIZATION, AND HYDROXYLATION.  
 RC TISSUE=Head;  
 RX MEDLINE=94342269; PubMed=8063725;  
 RA Dube H., Johnsen A.H., Scott A.G., East P., Thorpe A.;  
 RT "[Hyp3]Met-callistostatin. Identification and biological properties of  
 RT a novel neuropeptide from the blowfly Calliphora vomitoria.";  
 RL J. Biol. Chem. 269:21059-21066(1994).  
 CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR AND PLAY  
 CC A ROLE IN THE IMPREGATION OF INFORMATION WITHIN THE BRAIN. MAY BE  
 CC INVOLVED IN THE CONTROL OF VISCERAL MUSCLES DUE TO ITS ABILITY TO  
 CC BEHAVE AS POTENT INHIBITORS OF PERISTALTIC MOVEMENTS. MAY ALSO  
 CC FULFILL A NEUROHORMONAL ROLE ON MUSCLES OF THE GUT AND HEART.  
 CC -1- TISSUE SPECIFICITY: NEURONS WITHIN BRAIN AND ABDOMINAL GANGLION.  
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 DR PIR: E47393; E47393.  
 KM Neuropeptide; Amidation; Hydroxylation.  
 FT MOD\_RES 3 3  
 FT MOD\_RES 8 8 HYDROXYLATION (IN 20% OF THE PEPTIDES).  
 FT SEQUENCE 8 AA; 883 MW; 709879CABBA477768 CRC64;

Query Match 27.9%; Score 12; DB 1; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DFT 3  
 DB 5 DFT 6

RESULT 13  
 CCKN\_MACEU STANDARD; PRT; 8 AA.  
 ID CCKN\_MACEU  
 AC P30369;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Cholecystokinin (CCK).  
 GN CCK.

OS Macropus eugenii (Tamar wallaby), and  
OS Dasypus viverrinus (Southeastern quoll).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.  
OX NCBI\_TaxID=9315, 9279;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Brain;  
RX MEDLINE=88234141; PubMed=3375140;  
RA Fan Z.W., Eng J., Shaw G., Yalow R.S.;  
RT "Cholecystokinin octapeptide purified from brains of Australian  
marsupials";  
RT Peptides 9:429-431(1988).  
CC -1- FUNCTION: THIS PEPTIDE HORMONE INDUCES GALL BLADDER CONTRACTION  
AND THE RELEASE OF PANCREATIC ENZYMES IN THE GUT. ITS FUNCTION  
IN THE BRAIN IS NOT CLEAR.  
CC -1- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.  
DR PIR: A43001; A43001.  
DR PIR: P00012; P00012.  
DR InterPro: IPR001651; Gastrin.  
DR ProSite: PS00259; GASTRIN; 1.  
KW Amidation; Sulfation; Hormone.  
FT MOD\_RES 2 8  
FT MOD\_RES 8 8  
SQ SEQUENCE 8 AA; 1064 MW; DCA6837876B5A CRC64;

Query Match 27.9%; Score 12; DB 1; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1e+05; 0; Indels 0; Gaps 0;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 DF 3  
DB 7 DF 8

RESULT 14  
COXG\_RAT  
ID COXG\_RAT STANDARD; PRT; 8 AA.  
AC P80430;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Cytochrome c oxidase polypeptide VIb (EC 1.9.3.1) (AED) (Fragment).  
GN COX6B.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE.  
RC STRAIN=MSTAR; TISSUE=Liver;  
RX MEDLINE=95324529; PubMed=7601105;  
RA Schaeffer H., Noack H., Halanek W., Brandt U., von Jagow G.;  
RT "Cytochrome-c oxidase in developing rat heart. Enzymic properties and  
amino-terminal sequences suggest identity of the fetal heart and the  
adult liver isozyme";  
RT Eur. J. Biochem. 230:235-241(1995).  
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE  
CHAINS OF CYTOCHROME C OXIDASE, THE TERMINAL OXIDASE IN  
MITOCHONDRIAL ELECTRON TRANSPORT.  
CC -1- FUNCTION: THIS PROTEIN MAY BE ONE OF THE HEME-BINDING SUBUNITS OF  
THE OXIDASE.  
CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) -> 4 ferricytochrome  
c + 2 H(2)O.  
KW Oxidoreductase; Mitochondrion.  
FT NON\_TER 1 1  
FT NON\_TER 8 8  
SQ SEQUENCE 8 AA; 1039 MW; 8101E9CAA73AE456 CRC64;

Query Match 27.9%; Score 12; DB 1; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 DF 3  
DB 5 DF 6

RESULT 15  
FAR8\_CALVO  
ID FAR8\_CALVO STANDARD; PRT; 8 AA.  
AC P41863;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Calliphorinamide 8.  
OS Calliphora vomitoria (Blue blowfly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Oestrioidae; Calliphoridae; Calliphora.  
OX NCBI\_TaxID=27454;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Thoracic ganglion;  
RX MEDLINE=92196111; PubMed=1549595;  
RA Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,  
RA Rehfeld J.F., Thorpe A.;  
RT "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2  
RT neuropeptides (designated calliphorinamides) from the blowfly  
RT Calliphora vomitoria";  
RT Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).  
CC -1- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)  
FAMILY.  
CC PIR: H41978; H41978.  
DR Neuropeptide; Amidation.  
KW MOD\_RES 8 8  
FT MOD\_RES 8 8  
SQ SEQUENCE 8 AA; 957 MW; 72D40699CAA44DD8 CRC64;

Query Match 27.9%; Score 12; DB 1; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1e+05; 0; Indels 0; Gaps 0;  
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 DF 3  
DB 4 DF 5  
Search completed: July 15, 2002, 13:46:33  
Job time: 192 sec

4.

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:43:01 ; Search time 23.8 Seconds  
(without alignments)  
65.418 Million cell updates/sec

Title: US-09-712-819A-7  
Sequence: 1 TDFLTITSS 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 648

Minimum DB seq length: 0  
Maximum DB seq length: 9

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SPREMBL\_19:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protocist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	39.5	8	2	P77556
2	17	39.5	8	2	P77556
3	17	39.5	8	4	Q15902
4	17	39.5	8	7	Q29810
5	16	37.2	8	4	Q9UWH9
6	16	37.2	8	12	Q9IU21
7	16	37.2	8	12	Q9IU21
8	16	37.2	8	12	Q9IU21
9	15	34.9	8	4	Q9UD24
10	15	34.9	8	6	Q9UW78
11	15	34.9	9	2	Q9R635
12	15	34.9	9	11	Q9QVH9
13	15	34.9	9	13	P83057
14	14	32.6	9	2	Q31363
15	14	32.6	9	8	Q94VE1
16	14	32.6	9	8	Q94VD8

17	14	32.6	9	15	085723	085723 simian sarc
18	13	30.2	8	7	Q29801	Q29801 homo sapien
19	13	30.2	8	7	Q29797	Q29797 homo sapien
20	13	30.2	8	7	Q29812	Q29812 homo sapien
21	13	30.2	9	8	Q94V10	Q94V10 varanus g1g
22	13	30.2	9	11	Q92012	Q92012 mus muscula
23	13	30.2	9	12	Q91LX8	Q91LX8 retroperito
24	12	27.9	5	13	P82070	P82070 illoria rub
25	12	27.9	7	2	Q34028	Q34028 sphingomona
26	12	27.9	8	2	Q56246	Q56246 thermophill
27	12	27.9	8	2	Q47273	Q47273 escherichia
28	12	27.9	8	3	P87225	P87225 saccharomyc
29	12	27.9	8	3	Q60773	Q60773 homo sapien
30	12	27.9	8	4	Q9N6M5	Q9N6M5 toxoplasma
31	12	27.9	8	8	P92422	P92422 psathyrosta
32	12	27.9	8	8	P92373	P92373 haynaldia v
33	12	27.9	8	8	P93985	P93985 aegilops co
34	12	27.9	8	8	P92404	P92404 lophopyrum
35	12	27.9	8	8	P92426	P92426 pseudoroegn
36	12	27.9	8	8	P93973	P93973 eremopyrum
37	12	27.9	8	8	P93970	P93970 eremopyrum
38	12	27.9	8	8	P92388	P92388 hemaridia p
39	12	27.9	8	8	P92428	P92428 peridictyon
40	12	27.9	8	8	P92391	P92391 heteranthei
41	12	27.9	8	8	P92227	P92227 crithopsis
42	12	27.9	8	8	P93963	P93963 psathyrosta
43	12	27.9	8	8	P93961	P93961 psathyrosta
44	12	27.9	8	8	P92215	P92215 amblyopyrum
45	12	27.9	8	8	P93981	P93981 crithodium

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	8 AA.
P77556	AC	P77556:		
DR	01-FEB-1997 (TREMUREL.. 02, Created)			
DR	01-FEB-1997 (TREMUREL.. 02, Last sequence update)			
DR	01-JAN-1999 (TREMUREL.. 09, Last annotation update)			
DE	TRAY (FRAGMENT).			
GN	TRAY.			
OS	Escherichia coli.			
OC	Plasmid IncFII R1.			
OC	Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID-562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-ECOR11;			
RX	MEDLINE-96400908; PubMed-8807284;			
RA	Boyd E.F., Hill C.W., Rich S.M., Harcl D.L.;			
RT	*Mosaic structure of plasmids from natural populations of Escherichia coli*;			
RL	Genetics 143:1091-1100(1996).			
DR	EMBL; U50661; AAC44245.1; -			
DR	EMBL; U50650; AAC44236.1; -			
DR	EMBL; U50651; AAC44235.1; -			
DR	EMBL; U50652; AAC44236.1; -			
DR	EMBL; U50653; AAC44237.1; -			
DR	EMBL; U50654; AAC44238.1; -			
DR	EMBL; U50655; AAC44239.1; -			
DR	EMBL; U50656; AAC44240.1; -			
DR	EMBL; U50657; AAC44241.1; -			
DR	EMBL; U50658; AAC44242.1; -			
DR	EMBL; U50659; AAC44243.1; -			
DR	EMBL; U50660; AAC44244.1; -			
KW	Plasmid.			
FT	NON-TER			
SO	SEQUENCE 8 AA; 834 MW; D335A5B0544735A1 CRC64;			

Query Match 39.5%; Score 17; DB 2; Length 8;  
 Best Local Similarity 66.7%; Pred. No. 5.6e+05;  
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 TTTISS 9  
 Db 2 SLTISS 7

RESULT 2

P72221

PRELIMINARY; PRT; 8 AA.

ID P72221

01-FEB-1997 (TREMBlrel. 02, Created)

DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)

01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE ALGINATE LYASE (FRAGMENT).

ALY.

GN

Pseudomonas sp. (strain OS-ALG-9).

OC Bacteria; Proteobacteria.

NCBI\_TaxID=86038;

RM

SEQUENCE FROM N.A.

RP

STRAIN-OS-ALG-9.

RA Fujiyama K.;

Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.

RL

SEQUENCE FROM N.A.

RC

MEDLINE=93329366; PubMed=8336113;

RA Maki H., Maki A., Fujiyama K., Kimoshita S., Yoshida T.;

"Cloning, sequence analysis and expression in Escherichia coli of a

RT gene encoding an alginase from Pseudomonas sp. OS-ALG-9.";

J. Gen. Microbiol. 139:987-993(1993).

DR EMBL: D38469; BAA21704.1; -.

LYase.

KM

NON\_TER

FT

SEQUENCE

8 AA; 841 MW; 461DDCC5A5B041BB CRC64;

OY

6 TISS 9

Db

2 TISS 5

Query Match 39.5%; Score 17; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3

Q15902

PRELIMINARY; PRT; 8 AA.

ID Q15902

01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE (CLONE XP7E7A) (FRAGMENT).

Homo sapiens (Human).

OS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI\_TaxID=9606;

OX

SEQUENCE FROM N.A.

RP

TISSUE=PLACENTA;

RA Lee C.-C., Yezland A., Wehnert M., Bailey J., Couch L., Xiong M.,

Caskey C.T.H.;

RT "Isolation of chromosome-specific genes by reciprocal probing of

arrayed cDNAs and cosmid libraries.";

RL Hum. Mol. Genet. 0:0-0(1995).

EMBL: L33081; AAA73892.1; -.

FT

NON\_TER

SEQUENCE

7 AA; 684 MW; 71B5A5A5A2D1ABD0 CRC64;

FT NON\_TER 8 8  
 SQ SEQUENCE 8 AA; 931 MW; 83D699CAB1B2C9 CRC64;

Query Match 39.5%; Score 17; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TDF 3  
 Db 4 TDF 6

RESULT 4

Q29810

PRELIMINARY; PRT; 8 AA.

ID Q29810

01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE HLA CLASS II DR-BETA CHAIN (FRAGMENT).

HLA-DRB5\*15.

GN

Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX

NCBI\_TaxID=9606;

RM

SEQUENCE FROM N.A.

RP

MEDLINE=96175156; PubMed=9110934;

RA Svensson A.C., Setterblad N., Pihlgren U., Rask L., Andersson G.;

"Evolutionary relationship between different human major

RT histocompatibility complex HLA-DR haplotypes.";

Immunogenetics 43:304-314(1996).

DR EMBL: X88792; CAA61271.1; -.

NON\_TER

FT

SEQUENCE

8 AA; 909 MW; 6DC6C1A9CA731A8 CRC64;

OY

2 DPT 4

Db

4 DPT 6

Query Match 39.5%; Score 17; DB 7; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

P70804

PRELIMINARY; PRT; 7 AA.

ID P70804

01-FEB-1997 (TREMBlrel. 02, Created)

DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)

01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE ALGT PROTEIN (FRAGMENT).

ALGT.

OS

Azotobacter vinelandii.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

Azotobacter.

OX

NCBI\_TaxID=354;

RM

SEQUENCE FROM N.A.

RP

STRAIN=E;

RA MEDLINE=96427318; PubMed=8830682;

Rehm B.H.A., Ertesvag H., Valla S.;

RT "A new Azotobacter vinelandii mannuron C-5-epimerase gene (alg5) is

part of an alg gene cluster physically organized in a manner similar

RL to that in Pseudomonas aeruginosa.";

J. Bacteriol. 178:5884-5893(1996).

DR EMBL: X87973; CAA61230.1; -.

NON\_TER

SEQUENCE

7 AA; 684 MW; 71B5A5A5A2D1ABD0 CRC64;

```

Query Match          37.2%; Score 16; DB 2; Length 7;
Best Local Similarity 75.0%; Pred. NO. 5.6e+05;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 6 TISS 9
    1:11
Db 2 TVSS 5

RESULT 6
Q9UMH9 PRELIMINARY; PRT; 8 AA.
ID Q9UMH9;
AC Q9UMH9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE RHCE PROTEIN (FRAGMENT).
GN RHCE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE-BLOOD;
RA Matsui G., Cherif-Zahar B., Mouto I., Cartton J.P.;
RT "Characterization of the recombination hot spot involved in the
RT genomic rearrangement leading to the hybrid D-CE-D gene in the DVI
RT phenotype."
RL Am. J. Hum. Genet. 60:808-817(1997).
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE-BLOOD;
RA MEDLINE=90349591; PubMed=1696722;
RA Cherif-Zahar B., Bloy C., le Van Kim C., Blanchard D., Bailly P.,
RA Hermand P., Salmon C., Cartton J.-P., Colin Y.;
RT "Molecular cloning and protein structure of a human blood group Rh
RT polypeptide."
RL Proc. Natl. Acad. Sci. U.S.A. 87:6243-6247(1990).
DR EMBL; Z97030; CAB09726.1; -.
FT NON_TER 1 1
FT NON_TER 8 8
SO SEQUENCE 8 AA; 1049 MW; C007244691F85AB1 CRC64;

Query Match          37.2%; Score 16; DB 4; Length 8;
Best Local Similarity 28.6%; Pred. NO. 5.6e+05;
Matches 2; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TDFTLFI 7
    1:11
Db 1 TDFHML 7

RESULT 7
Q91U21 PRELIMINARY; PRT; 8 AA.
ID Q91U21;
AC Q91U21;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NUCLEAR EXPORT PROTEIN NS2 (FRAGMENT).
GN NS.
OS Influenza A virus (A/Hong Kong/491/97(H5N1)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX NCBI_TaxID=155223;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=A/HONG KONG/491/97;

```

```

RA Shaw M.W., Cooper L.A., Xu X., Thompson W.W., Kraus S.L., Guan Y.,
RA Zhou N.N., Klimov A., Cox N.J., Webster R.G., Lim W., Shortridge K.F.,
RA Subbarao K.;
RT "Avian Influenza A H5N1 and H9N2 viruses bearing a specific
RT constellation of nonglycoprotein genes caused illness in humans."
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF256191; AAK49324.1; -.
FT NON_TER 1 1
FT NON_TER 8 8
SO SEQUENCE 8 AA; 869 MW; D9D9D5A5A2D1A455 CRC64;

Query Match          37.2%; Score 16; DB 12; Length 8;
Best Local Similarity 75.0%; Pred. NO. 5.6e+05;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 6 TISS 9
    1:11
Db 3 TVSS 6

RESULT 8
Q91U19 PRELIMINARY; PRT; 8 AA.
ID Q91U19;
AC Q91U19;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NUCLEAR EXPORT PROTEIN NS2 (FRAGMENT).
GN NS.
OS Influenza A virus (A/Hong Kong/503/97(H5N1)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX NCBI_TaxID=155224;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=A/HONG KONG/503/97;
RA Shaw M.W., Cooper L.A., Xu X., Thompson W.W., Kraus S.L., Guan Y.,
RA Zhou N.N., Klimov A., Cox N.J., Webster R.G., Lim W., Shortridge K.F.,
RA Subbarao K.;
RT "Avian Influenza A H5N1 and H9N2 viruses bearing a specific
RT constellation of nonglycoprotein genes caused illness in humans."
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF256192; AAK49326.1; -.
FT NON_TER 1 1
FT NON_TER 8 8
SO SEQUENCE 8 AA; 869 MW; D9D9D5A5A2D1A455 CRC64;

Query Match          37.2%; Score 16; DB 12; Length 8;
Best Local Similarity 75.0%; Pred. NO. 5.6e+05;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 6 TISS 9
    1:11
Db 3 TVSS 6

RESULT 9
Q9UD24 PRELIMINARY; PRT; 8 AA.
ID Q9UD24;
AC Q9UD24;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RHD PROTEIN (FRAGMENT).
GN RHD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.

```

RC TISSUE-BLOOD;  
 RA MEDLINE-97260406; PubMed-9105526;  
 RA Matsui G., Cherif-Zahar B., Moura I., Cartton J.P.;  
 RT "Characterization of the recombination hot spot involved in the  
 RT genomic rearrangement leading to the hybrid D-CE-D gene in the DVI  
 RT phenotype.";  
 RL Am. J. Hum. Genet. 60:808-817(1997).  
 RN (12)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BLOOD;  
 RX MEDLINE-93066356; PubMed-1438298;  
 RA Le Van Kim C., Moura I., Cherif-Zahar B., Reynal V., Chertier C.,  
 RT "Molecular cloning and primary structure of the human blood group Rh  
 RT polypeptide.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10925-10929(1992).  
 DR EMBL; Z97031; CAB09727.1; -  
 FT NON\_TER 1 1  
 FT NON\_TER 8 8  
 SQ SEQUENCE 8 AA; 1042 MW; D296944691FB5AB1 CRC64;

Query Match 34.9%; Score 15; DB 4; Length 8;  
 Best Local Similarity 28.6%; Pred. No. 5.6e+05;  
 Matches 2; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 TDFTLTI 7  
 DB 1 TDYHNM 7

RESULT 10  
 O9TT78 PRELIMINARY; PRT; 8 AA.  
 AC O9TT78;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE TRIMIDYLATE SYNTHASE (FRAGMENT).  
 GN TS.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 RX NCB1\_TaxID=9615;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-21015404; PubMed-11130975;  
 RA Brouillette J.A., Andrew J.R., Venta P.J.;  
 RT "Estimate of nucleotide diversity in dogs with a pool-and-sequence  
 RT method.";  
 RL Mamm. Genome 11:1079-1086(2000).  
 DR EMBL; AF202073; AAF20918.1; -  
 FT NON\_TER 1 1  
 FT NON\_TER 8 8  
 SQ SEQUENCE 8 AA; 899 MW; 6731A1E059CA867 CRC64;

Query Match 34.9%; Score 15; DB 6; Length 8;  
 Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
 Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 DFTLTI 7  
 DB 3 DFTLTI 8

RESULT 11  
 O9R635 PRELIMINARY; PRT; 9 AA.  
 AC O9R635;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)

DE MAJOR OUTER MEMBRANE PROTEIN VARIABLE DOMAIN IV, MONO VD IV.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 RX NCB1\_TaxID=813;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-92040090; PubMed-1718870;  
 RA Peterson E.M., Cheng X., Markoff B.A., Fielder T.J., de la Maza L.M.;  
 RT "Functional and structural mapping of Chlamydia trachomatis species-  
 RT specific major outer membrane protein epitopes by use of neutralizing  
 RT monoclonal antibodies.";  
 RL Infect. Immun. 59:4147-4153(1991).  
 SQ SEQUENCE 9 AA; 976 MW; 9C61B041B7645361 CRC64;

Query Match 34.9%; Score 15; DB 2; Length 9;  
 Best Local Similarity 50.0%; Pred. No. 5.6e+05;  
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 TDFTLTI 8  
 DB 1 TTWNPTIS 8

RESULT 12  
 O9OVH9 PRELIMINARY; PRT; 9 AA.  
 AC O9OVH9;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
 DE SUPERINDUCIBLE PROTEIN 24, SIP24-CYCLOPHILIN HOMOLOG, PEAK C.  
 OS Mus sp.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RX NCB1\_TaxID=10095;  
 RN (1)  
 RP SEQUENCE.  
 RX MEDLINE-92031730; PubMed-1932134;  
 RA Davis T.R., Tabatabai L., Bruns K., Hamilton R.T., Nilsen-Hamilton M.;  
 RT "Basic fibroblast growth factor induces 373 fibroblasts to synthesize  
 RT and secrete a cyclophilin-like protein and beta 2-microglobulin.";  
 RL Biochim. Biophys. Acta 1095:145-152(1991).  
 SQ SEQUENCE 9 AA; 963 MW; D6D2731A9D5087D CRC64;

Query Match 34.9%; Score 15; DB 11; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 FTL 5  
 DB 5 FTL 7

RESULT 13  
 P83057 PRELIMINARY; PRT; 9 AA.  
 AC P83057;  
 DT 01-OCT-2001 (Tremblrel. 18, Created)  
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)  
 DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)  
 DE BRADYKININ-LIKE PEPTIDE ([VAL1,THR3,THR6]BRADYKININ).  
 OS Bombina variegata (yellow-bellied toad).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Archeobatrachia; Bombinatoridae; Bombina.  
 RX NCB1\_TaxID=8348;  
 RN (1)  
 RP SEQUENCE, SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.  
 RC TISSUE=SKIN SECRETION;  
 RA Chen T.B., Orr D.F., Bjorson A.J., McClean S., Rao P.F., Shaw C.;  
 RT "Cloning and post-translational processing of frog skin kininogens.";  
 RL Submitted (JUL-2001) to the SWISS-PROT data bank.



CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- TISSUE SPECIFICITY: SKIN.  
SQ SEQUENCE 9 AA: 1021 MW: 36870771A9C861B7 CRC64:

Query Match 34.9%; Score 15; DB 13; Length 9;  
Best Local Similarity 75.0%; Pred. No. 5.6e+05;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TDT 4  
111  
Db 3 TSFT 6

Search completed: July 15, 2002, 13:46:16  
Job time: 195 sec

Query Match 32.6%; Score 14; DB 8; Length 9;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TLT 6  
111  
Db 2 TLT 4

RESULT 14  
O31363 PRELIMINARY; PTR; 9 AA.  
AC O31363;  
DT 01-JAN-1998 (TREMblrel. 05, Created)  
DT 01-JAN-1998 (TREMblrel. 05, last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
DE OUTER SURFACE PROTEIN C (FRAGMENT).  
GN OSPC.  
OS Borrelia garinii.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_TaxID-29519;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PRI.  
RX MEDLINE-97426044; Pubmed-9282748;  
RA Tilly K., Casjens S., Stevenson B., Bono J.L., Samuels D.S., Hogan D., Rosa P.;  
RT "The Borrelia burgdorferi circular plasmid cp26: conservation of plasmid structure and targeted inactivation of the ospC gene.";  
RL MOL. Microbiol. 25:361-374(1997).  
DR EMBL: U93699; AAC4553.1; -;  
FT NON\_TER 9  
SQ SEQUENCE 9 AA: 1019 MW: 4864C1A731A4433 CRC64:

Query Match 32.6%; Score 14; DB 2; Length 9;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TLT 6  
111  
Db 5 TLT 7

RESULT 15  
Q94VE1 PRELIMINARY; PTR; 9 AA.  
AC Q94VE1;  
DT 01-DEC-2001 (TREMblrel. 19, Created)  
DT 01-DEC-2001 (TREMblrel. 19, last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
DE CYTOCHROME C OXIDASE SUBUNIT I (FRAGMENT).  
GN COI.  
OS Varanus meriensis.  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Lepidosauria; Squamata; Scleroglossae; Anguimorpha; Varanidae; Varanus.  
OX NCBI\_TaxID-62044;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ast J.C.;  
RT "Mitochondrial DNA evidence and evolution in Varanoidae (Squamata).";  
RL Cladistics 17:0-0(2001).  
DR EMBL: AF407512; AAL10090.1; -;  
KW Mitochondrion.  
FT NON\_TER 9  
SQ SEQUENCE 9 AA: 1154 MW: 9E80C7336411A731 CRC64:

10

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:39:30 ; Search time 28.73 Seconds  
(without alignments)

34.795 Million cell updates/sec

Title: US-09-712-819A-7

Perfect score: 43

Sequence: 1 TDFLTITSS 9

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 102553

Minimum DB seq length: 0  
Maximum DB seq length: 9

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*

1: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgcdata/hold-genseq/genseqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	79.1	7	20	AAV40738
2	34	79.1	7	21	AAB30076
3	31	72.1	7	20	AAV40737
4	31	72.1	7	21	AAB30075
5	30	69.8	7	20	AAV40736
6	30	69.8	7	21	AAB30074
7	25	58.1	8	18	AAV19840
8	23	53.5	8	16	AAV77190
9	22	51.2	9	22	AAB56069
10	22	51.2	9	22	AAU28452
11	22	51.2	9	22	AAU28098

12	21	48.8	7	20	AAV40735
13	21	48.8	7	20	AAV40736
14	21	48.8	7	21	AAB30073
15	21	48.8	9	16	AAB79659
16	21	48.8	9	22	AAV79658
17	21	48.8	9	22	AAV79659
18	20	46.5	6	19	AAV31467
19	20	46.5	8	14	AAV34415
20	20	46.5	8	16	AAV34415
21	20	46.5	8	16	AAV34415
22	20	46.5	8	22	AAU08941
23	20	46.5	9	15	AAV67110
24	20	46.5	9	15	AAV55716
25	20	46.5	9	15	AAV37078
26	20	46.5	9	21	AAV64433
27	20	46.5	9	22	AAV50029
28	19	44.2	5	9	AAV82200
29	19	44.2	6	20	AAV06466
30	19	44.2	6	21	AAV77708
31	19	44.2	6	22	AAV13076
32	19	44.2	6	22	AAV51422
33	19	44.2	6	22	AAV97355
34	19	44.2	7	22	AAV59859
35	19	44.2	7	21	AAV2584
36	19	44.2	8	9	AAV82196
37	19	44.2	8	16	AAV6167
38	19	44.2	8	21	AAV01851
39	19	44.2	8	21	AAV61944
40	19	44.2	8	21	AAV63487
41	19	44.2	9	15	AAV37969
42	19	44.2	9	16	AAV78826
43	19	44.2	9	18	AAV00868
44	19	44.2	9	19	AAV21455
45	19	44.2	9	19	AAV77115

#### ALIGNMENTS

RESULT 1	
AAV40738	standard; peptide: 7 AA.
AAV40738:	
01-DEC-1999	(first entry)
XX	S4 derivative #12, beta strand of scaffold protein structure.
XX	Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;
XX	tumour; chemotherapeutic agent.
XX	Synthetic.
OS	
XX	
PV	EP947582-A1.
XX	
PD	06-OCT-1999.
XX	
PF	31-MAR-1998; 98BP-0870065.
XX	
PR	31-MAR-1998; 98BP-0870065.
XX	
PA	(INNO-) INNOGENETICS NV.
XX	
PI	Desmet J, Hufion S, Hoogenboom H, Sablon E;
XX	
DR	WPI: 1999-542958/46.
XX	
PT	New scaffold protein, useful for stabilizing antigens used as vaccines
PS	
XX	Disclosure; Page 6; 105pp; English.

CC Sequences AAY40727-Y40748 are functionally equivalent derivatives of the  
CC S4 peptide (AAY40607) which forms part of a scaffold protein. S4 is a  
CC beta strand peptide which forms part of a beta sheet. Peptides  
CC (AAY40601-Y40609) together form a single-chain scaffold protein which  
CC contains at least 1 disulfide bond, contains less than 10% alpha helix  
CC and contains at least 6 beta-strands. The scaffold protein is constructed  
CC of beta strands S1-S6, and may also include beta strands A1-A3, or any  
CC functionally equivalent derivative of these sequences. The beta strands  
CC form two beta sheets S1/S4/S3 and S6/S5/S2 with each strand connected to  
CC the next by hydrogen bonds, which generate a beta sandwich architecture.  
CC If the additional beta strands A1-A3 are included in the structure the  
CC scaffold is constructed of two beta sheets, with the structures  
CC A1/S1/S4/S3 and S6/S5/S2/A2/A3. The beta strands are connected to each  
CC other via amino acid loops, where at least one of the loops binds to a  
CC receptor or antigen. The scaffold protein is used to stabilize antigens  
CC or whole proteins such as receptors, or their fragments. It may be used  
CC to bind two separate molecules. For example, one surface of the scaffold  
CC may be bound to a protein which binds to a tumour antigen. This will  
CC target the complex to tumour cells. Another surface may be bound to a  
CC cytotoxic molecule or an autoimmune antibody which may then kill the  
CC tumour cells. Therefore the scaffold protein may be used to target  
CC chemotherapeutic agents to specific cells. It may also be used to  
CC stabilize individual peptides in a peptide library and may be used in  
CC diagnostic techniques, and to stabilize antigens used as vaccines.

SO Sequence 7 AA:

Query Match 79.1%; Score 34; DB 20; Length 7;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DFTLTIS 8  
|||||||  
Db 1 dftltis 7

RESULT 2

AAB30076  
ID AAB30076 standard; Peptide: 7 AA.

XX AAB30076;

DT 09-FEB-2001 (first entry)

XX Scaffold protein SCA S4 peptide SEQ ID NO: 137.

DE Human: CTLA-4; scaffold protein; antigen-binding; receptor-binding;  
XX SCA domain; cancer; thrombosis; osteoporosis; rheumatoid arthritis;  
KW diabetic retinopathy; atherosclerosis.

XX Synthetic.

OS

XX WO200060070-A1.

XX 12-OCT-2000.

XX 01-APR-1999; 99WO-EP02283.

XX 01-APR-1999; 99WO-EP02283.

XX (INNO-) INNOGENETICS NV.

XX Desmet J, Hufion S, Hoogenboom H, Sablon E;

XX WPI; 2000-665002/64.

XX Scaffold composed of single-chain polypeptide having beta sandwich  
PT architecture carrying new and randomized peptide sequences useful as  
PT supporting framework and carrying antigen- or receptor binding  
XX fragments

PS Disclosure; Page 15; 68pp; English.

XX The present invention is concerned with producing scaffold proteins  
CC based upon the human CTLA-4 SCA domain. These scaffold proteins can be  
CC used as a scaffold to bind antigen- or receptor-binding fragments. These  
CC can be used in the treatment of diseases such as cancer,  
CC atherosclerosis, thrombosis, osteoporosis, rheumatoid arthritis and  
CC production of the proteins of the invention.

SO Sequence 7 AA:

Query Match 79.1%; Score 34; DB 21; Length 7;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 DFTLTIS 8  
|||||||  
Db 1 dftltis 7

RESULT 3

AAY40737  
ID AAY40737 standard; peptide: 7 AA.

XX AAY40737;

DT 01-DEC-1999 (first entry)

XX S4 derivative #11, beta strand of scaffold protein structure.

DE Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;  
KW tumour; chemotherapeutic agent.

XX Synthetic.

XX EP947582-A1.

XX 06-OCT-1999.

XX 31-MAR-1998; 98EP-0870065.

XX 31-MAR-1998; 98EP-0870065.

XX (INNO-) INNOGENETICS NV.

XX Desmet J, Hufion S, Hoogenboom H, Sablon E;

XX WPI; 1999-542958/46.

XX New scaffold protein, useful for stabilizing antigens used as vaccines  
PT  
PT  
PT  
PS Disclosure; Page 6; 105pp; English.

CC Sequences AAY40727-Y40748 are functionally equivalent derivatives of the  
CC S4 peptide (AAY40607) which forms part of a scaffold protein. S4 is a  
CC beta strand peptide which forms part of a beta sheet. Peptides  
CC (AAY40601-Y40609) together form a single-chain scaffold protein which  
CC contains at least 1 disulfide bond, contains less than 10% alpha helix  
CC and contains at least 6 beta-strands. The scaffold protein is constructed  
CC of beta strands S1-S6, and may also include beta strands A1-A3, or any  
CC functionally equivalent derivative of these sequences. The beta strands  
CC form two beta sheets S1/S4/S3 and S6/S5/S2 with each strand connected to  
CC the next by hydrogen bonds, which generate a beta sandwich architecture.  
CC If the additional beta strands A1-A3 are included in the structure the  
CC scaffold is constructed of two beta sheets, with the structures  
CC A1/S1/S4/S3 and S6/S5/S2/A2/A3. The beta strands are connected to each  
CC other via amino acid loops, where at least one of the loops binds to a  
CC receptor or antigen. The scaffold protein is used to stabilize antigens  
CC or whole proteins such as receptors, or their fragments. It may be used  
CC to bind two separate molecules. For example, one surface of the scaffold  
CC may be bound to a protein which binds to a tumour antigen. This will

target, the complex to tumour cells. Another surface may be bound to a cytotoxic molecule or an autoimmune antibody which may then kill the tumour cells. Therefore the scaffold protein may be used to target chemotherapeutic agents to specific cells. It may also be used to stabilize individual peptides in a peptide library and may be used in diagnostic techniques, and to stabilize antigens used as vaccines.

XX  
SQ Sequence 7 AA;

Query Match 72.1%; Score 31; DB 20; Length 7;  
Best Local Similarity 85.7%; Pred. No. 6.4e+05;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 DFLTIS 8  
1:|||||

DB 1 dyflits 7

RESULT 4

AAB30075  
ID AAB30075 standard; Peptide; 7 AA.

XX  
AC AAB30075;

XX  
DT 09-FEB-2001 (first entry)

XX  
DE Scaffold protein SCA S4 peptide SEQ ID NO: 136.

XX  
KW Human; CTLA-4; scaffold protein; antigen-binding; receptor-binding;

XX  
KM SCA domain; cancer; thrombosis; osteoporosis; rheumatoid arthritis;

XX  
PT diabetic retinopathy; atherosclerosis.

XX  
OS Synthetic.

XX  
PN WO200060070-A1.

XX  
PD 12-OCT-2000.

XX  
PF 01-APR-1999; 99WO-EP02283.

XX  
PR 01-APR-1999; 99WO-EP02283.

XX  
PA (INNO-) INNOGENETICS NV.

XX  
PI Desmet J, Hufton S, Hoogenboom H, Sablon E;

XX  
DR WPI: 2000-665002/64.

XX  
PT Scaffold composed of single-chain polypeptide having beta sandwich

XX  
PF architecture carrying new and randomized peptide sequences useful as

XX  
PT supporting framework and carrying antigen- or receptor binding

XX  
PT fragments

XX  
PS Disclosure: Page 15; 68pp; English.

XX  
CC The present invention is concerned with producing scaffold proteins

XX  
CC based upon the human CTLA-4 SCA domain. These scaffold proteins can be

XX  
CC used as a scaffold to bind antigen- or receptor-binding fragments. These

XX  
CC can be used in the treatment of diseases such as cancer,

XX  
CC atherosclerosis, thrombosis, osteoporosis, rheumatoid arthritis and

XX  
CC diabetic retinopathy. Sequences AAB3930-B29939 were used in the

XX  
SQ Sequence 7 AA;

QY

1:|||||

DB 1 dyflits 7

RESULT 5

AAY40736  
ID AAY40736 standard; peptide; 7 AA.

XX  
AC AAY40736;

XX  
DT 01-DEC-1999 (first entry)

XX  
DE S4 derivative #10, beta strand of scaffold protein structure.

XX  
KW Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;

XX  
KM tumour; chemotherapeutic agent.

XX  
OS Synthetic.

XX  
PN EP947582-A1.

XX  
PD 06-OCT-1999.

XX  
PF 31-MAR-1998; 98EP-0870065.

XX  
PR 31-MAR-1998; 98EP-0870065.

XX  
PA (INNO-) INNOGENETICS NV.

XX  
PI Desmet J, Hufton S, Hoogenboom H, Sablon E;

XX  
DR WPI: 1999-542958/46.

XX  
PT New scaffold protein, useful for stabilizing antigens used as vaccines

XX  
PT -

XX  
PS Disclosure: Page 6; 105pp; English.

XX  
CC Sequences AAY40727-Y40748 are functionally equivalent derivatives of the

XX  
CC S4 peptide (AAY40607) which forms part of a scaffold protein. S4 is a

XX  
CC beta strand peptide which forms part of a beta sheet. Peptides

XX  
CC (AAY40601-Y40609) together form a single-chain scaffold protein which

XX  
CC contains at least 1 disulfide bond, contains less than 10% alpha helix

XX  
CC and contains at least 6 beta-strands. The scaffold protein is constructed

XX  
CC of beta strands S1-S6, and may also include beta strands A1-A3, or any

XX  
CC functionally equivalent derivative of these sequences. The beta strands

XX  
CC form two beta sheets S1/S4/S3 and S6/S5/S2 with each strand connected to

XX  
CC the next by hydrogen bonds, which generate a beta sandwich architecture.

XX  
CC If the additional beta strands A1-A3 are included in the structure the

XX  
CC scaffold is constructed of two beta sheets, with the structures

XX  
CC A1/S1/S4/S3 and S6/S5/S2/A2/A3. The beta strands are connected to each

XX  
CC other via amino acid loops, where at least one of the loops binds to a

XX  
CC receptor or antigen. The scaffold protein is used to stabilize antigens

XX  
CC or whole proteins such as receptors, or their fragments. It may be used

XX  
CC to bind two separate molecules. For example, one surface of the scaffold

XX  
CC may be bound to a protein which binds to a tumour antigen. This will

XX  
CC target the complex to tumour cells. Another surface may be bound to a

XX  
CC cytotoxic molecule or an autoimmune antibody which may then kill the

XX  
CC tumour cells. Therefore the scaffold protein may be used to target

XX  
CC chemotherapeutic agents to specific cells. It may also be used to

XX  
CC stabilize individual peptides in a peptide library and may be used in

XX  
CC diagnostic techniques, and to stabilize antigens used as vaccines.

XX  
SQ Sequence 7 AA;

QY

1:|||||

DB 1 dflits 7

Query Match 69.8%; Score 30; DB 20; Length 7;  
Best Local Similarity 85.7%; Pred. No. 6.4e+05;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 DFLTIS 8  
1:|||||

DB 1 dflits 7

```

RESULT 6
AAB30074
ID AAB30074 standard; Peptide; 7 AA.
XX
XX
AC AAB30074;
XX
XX
DT 09-FEB-2001 (first entry)
XX
DE Scaffold protein SCA 54 peptide SEQ ID NO: 135.
XX
XX
KW Human; CTLA-4; scaffold protein; antigen-binding; receptor-binding;
KW SCA domain; cancer; thrombosis; osteoporosis; rheumatoid arthritis;
KW diabetic retinopathy; atherosclerosis.
XX
OS Synthetic.
XX
XX
PN WO200060070-A1.
XX
PD 12-OCT-2000.
XX
XX
PE 01-APR-1999; 99WO-EP02283.
XX
XX
PR 01-APR-1999; 99WO-EP02283.
XX
XX
PA (INNO-) INNOGENETICS NV.
XX
PI Deemet J, Hufton S, Hoogenboom H, Sablon E;
XX
XX
DR WPI; 2000-665002/64.
XX
XX
PT Scaffold composed of single-chain polypeptide having beta sandwich
PT architecture carrying new and randomized peptide sequences useful as
PT supporting framework and carrying antigen- or receptor binding
PT fragments.
XX
XX
PS Disclosure; Page 15; 68pp; English.
XX
XX
CC The present invention is concerned with producing scaffold proteins
CC based upon the human CTLA-4 SCA domain. These scaffold proteins can be
CC used as a scaffold to bind antigen- or receptor-binding fragments. These
CC can be used in the treatment of diseases such as cancer,
CC atherosclerosis, thrombosis, osteoporosis, rheumatoid arthritis and
CC production of the proteins of the invention.
CC
XX
SQ Sequence 7 AA:

```

```

Query Match 69.8%; Score 30; DB 21; Length 7;
Best Local Similarity 85.7%; Pred. NO. 6.4e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 DFTLTIS 8
   |||||
   1 dftltis 7
Db 1 dftltis 7

RESULT 7
AAW19840
ID AAW19840 standard; Peptide; 8 AA.
XX
XX
AC AAW19840;
XX
XX
DT 26-JAN-1998 (first entry)
XX
XX
DE Chimeric adenovirus coat protein heparin binding motif.
XX
KW Adenovirus; vector; coat protein; gene therapy; gene transfer;
KW human; cancer; autoimmune disease; heart disease; infection;
KW heparin.
XX
OS Synthetic.

```

```

XX
XX
PN WO9720051-A2.
XX
XX
PD 05-JUN-1997.
XX
XX
PE 27-NOV-1996; 96WO-US19150.
XX
XX
PR 21-AUG-1996; 96US-0701124.
PR 28-NOV-1995; 95US-0563368.
PR 21-AUG-1996; 96US-0700846.
XX
XX
PA (GENV-) GENVEC INC.
XX
PI Brough DE, Kovesal I, Wickham TJ;
XX
XX
DR WPI; 1997-310606/28.
XX
XX
PT Adenoviral vectors containing chimeric coat protein - bind and enter
PT cells more efficiently, useful for gene therapy of e.g. cancer,
PT autoimmune diseases, etc.
XX
XX
PS Claim 7; Page 19; 121pp; English.
XX
XX
CC This peptide is used as a universal transfer vector (UTV) sequence
CC or as a spacer sequence in novel chimeric adenovirus coat proteins
CC (CP), especially chimeric fibre proteins. It is modeled after
CC fibronectin and exhibits heparin binding properties. Claimed
CC UTVs/spacers are given in AAW19810-11, AAW19813-25, AAW19827, AAW19829,
CC AAW19831-32 and AAW19834-43). Claimed chimeric CPs differ from the
CC wild-type CP by the introduction of the UTV and/or spacer at or
CC near the C-terminus or in an exposed loop. This impacts on the
CC chimeric CP the ability to bind to and enter cells by means of a
CC novel cell surface binding site. Recombinant vectors comprising
CC the chimeric CP are able to enter cells more efficiently than
CC vectors comprising wild-type CP, especially at lower m.o.i. They
CC are especially useful for gene therapy of e.g. cancers, genetic
CC disorders, pathogenic infections, heart disease or autoimmune
CC diseases.
XX
XX
SQ Sequence 8 AA:

```

```

Query Match 58.1%; Score 25; DB 18; Length 8;
Best Local Similarity 62.5%; Pred. NO. 6.4e+05;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 TDTLTIS 8
   |||||
   1 tdtltis 8
Db 1 tdtltis 8

RESULT 8
AAR77190
ID AAR77190 standard; Protein; 8 AA.
XX
XX
AC AAR77190;
XX
XX
DT 24-NOV-1995 (first entry)
XX
XX
DE Internal sequence of cytochrome P-450OX.
XX
KW Cytochrome; P-450ox; monooxygenase.
XX
OS Synthetic.
XX
XX
PN WO9516041-A.
XX
XX
PD 15-JUN-1995.
XX
XX
PR 28-NOV-1994; 94WO-EP03938.
XX
XX
PR 08-DEC-1993; 93EP-0810860.
XX

```

PA	(CIBA-) <sup>y</sup> CIBA GEIGY AG.
PA	(UYRO-) UNIV ROYAL VETERINARY & AGRIC.
XX	
PI	Halkier BA, Koch BM, Lindberg MÜLER B, Stibsen O;
PI	Lindberg MÖLLER B;
XX	
DR	WPI; 1995-224324/29.
XX	
PT	DNA encoding cytochrome P-450 mono:oxygenase(s) - for modifying
PT	biosynthetic production of cyanogenic glycoside(s) and
PT	glucosinolate(s) in crop plant(s) for increased pest resistance
PT	or improved nutritive value
PS	
PS	Example: Page 27; 73pp; English.
XX	
CC	Cytochrome P-450OX resembles the cytochrome P-450 reported to
CC	convert oximes to nitriles in rat liver microsomes. The mol. wt.
CC	of P-450OX as determined by SDS-PAGE is 51 kD. AA sequences are
CC	obtd. by Edman degradation. The internal polypeptides are obtd.
CC	by trypsin digestion of the purified protein and subsequent sepn.
CC	of peptides using reverse phase HPLC. The N-terminal sequence is
CC	given in AAR77188. A cytochrome with a mol. wt. of 51 kD and this
CC	N-terminal sequence is claimed. Internal peptides are given in
CC	AAR77189-R77192.
XX	
SQ	Sequence 8 AA;
XX	
Query Match	53.5%; Score 23; DB 16; Length 8;
Best Local Similarity	80.0%; Pred. No. 6.4e+05;
Matches 4;	Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY	2 DFTLT 6
Db	2 dftvt 6
XX	
RESULT 9	
ID	ABBS6069 standard; Peptide: 9 AA.
XX	
AC	ABBS6069;
XX	
DT	15-FEB-2002 (first entry)
XX	
DE	Vascular dementia-associated protein isoform (VPI) 269.
XX	
KM	Vascular Dementia; VD: VD-associated protein isoform; VPI; screening;
XX	diagnosis; prognosis; gene therapy.
XX	
OS	Homo sapiens.
XX	
FN	M0200169261-A2.
XX	
DD	20-SEP-2001.
XX	
PF	14-MAR-2001; 2001MC-GB01106.
XX	
PR	15-MAR-2000; 2000GB-0006285.
XX	
PR	24-NOV-2000; 2000GB-0028734.
XX	
PR	28-NOV-2000; 2000US-0724391.
XX	
PA	(OXFO-) OXFORD GLYCOSCIENTES UK LTD.
XX	
PT	Herath HMAG, Parekh RB, Rohlfé C;
XX	
DR	WPI; 2001-557937/62.
XX	
PT	Screening, diagnosis or prognosis of vascular dementia (VD), useful for
PT	determining stage of VD and monitoring the effect of VD therapy,
PT	comprises analysing body fluid by 2-dimensional electrophoresis for
PT	features correlated with VD -
XX	

PS Claim 6; Page 35; 151pp; English.

The invention relates to screening, diagnosis or prognosis of Vascular Dementia (VD) in a subject comprising analysing body fluid from the subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of features containing at least one chosen feature whose relative abundance correlates with the presence, absence, stage or severity of VD or predicts the onset or course of VD, especially detecting in a sample of cerebrospinal fluid (CSF) from the subject one of 223 VD-associated protein isoforms (VPIs) (ABBS5801-ABBS6295) as fully defined in the specification. Detecting VD-associated features and VPI is useful for the screening, diagnosis or prognosis of VD, for determining the stage or severity of VD, for identifying a subject at risk of VD or for monitoring the effect of therapy administered to a subject having VD. Nucleic acids encoding a VPI or inhibiting the function of a VPI are useful for the treatment of VD and for gene therapy.

Sequence 9 AA:

Query Match 51.2% Score 22; DB 22; Length 9;  
Best Local Similarity 80.0%; Prod No. 6.4e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps

OY 2 DEFULT 6  
||||:  
Db 2 dflts 6

RESULT 10  
AAU28452  
ID AAU28452 standard; Peptide; 9 AA.  
XX  
XX AAU28452:  
DY 03-JAN-2002 (first entry)  
XX  
XX DPI tryptic digest peptide #49.  
DE  
XX  
XX Human: depression associated protein isoform; tryptic digest peptide;  
KW DPI; cerebrospinal fluid; CSF; BAD; bipolar affective disorder;  
KM neuropsychiatric disorder; bipolar mood disorder; neuroleptic;  
RW manic-depressive illness; schizoaffective disorder.  
XX  
XX Homo sapiens.  
OS  
XX W0200162787-A1.  
PN  
XX 30-AUG-2001.  
PD  
XX  
PF 23-FEB-2001; 2001MO-GB00786.  
PP  
XX 24-FEB-2000; 2000GB-0004412.  
PR 08-DEC-2000; 2000GB-0030050.  
PDR 12-DEC-2000; 2000US-0254830.  
XXX  
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX  
XX Herach IMAC, Parekh RB, Rohlf C, Terrett JA, Tyson KL,  
PI  
XX WPI: 2001-570626/64.  
DR  
XX Novel nucleic acid encoding a protein associated with bipolar affective  
PT disorder, which is used for diagnosis, prophylaxis and therapy of  
PT neuropsychiatric disorders, such as bipolar affective disorder -  
XX  
XX Disclosure; page 31; 153pp; English.  
PS  
XX The present invention relates to the identification of depression  
CC associated protein isoforms (PPIs), particularly the tryptic digest  
CC peptides of these proteins. Some of the PPIs (AAU28404-AAU28625)  
CC described are decreased in the cerebrospinal fluid (CSF) of BAD  
CC (bipolar affective disorder) subjects, whilst other PPIs

CC (AAU28626-AAU28887) are increased in BAD subjects. Also described  
CC are peptide sequences identified from DPI-45 and DPI-213 and the  
CC nucleic acid sequences they are encoded by. The sequences of the  
CC invention are useful for clinical screening, diagnosis, prognosis,  
CC therapy and prophylaxis of neuropsychiatric disorders e.g. BAD (also  
CC known as bipolar mood disorder, BP), manic-depressive illnesses,  
CC attention deficit disorders, schizoaffective disorders, and unipolar  
CC affective disorders. The present sequence represents one of the DPI  
CC tryptic digest peptides of the present invention.  
XX  
SO Sequence 9 AA:  
  
Query Match 51.2%; Score 22; DB 22; Length 9;  
Best Local Similarity 80.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 DFTLT 6  
| | | | |  
Db 2 dftls 6  
  
RESULT 11  
AAU26098  
ID AAU26098 standard; Peptide: 9 AA.  
XX  
AC AAU26098;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Depression-Associated Protein Isoform DPI-9 #5.  
XX  
XX Human: Bipolar Affective Disorder; BAD; Depression-Associated feature;  
XX DF; Depression-Associated protein isoform; DPI; Cerebro-spinal fluid;  
XX CSF; antidepressant; anxiolytic; nootropic; tranquiliser; neuroleptic;  
XX attention deficient disorder; schizoaffective disorder;  
XX unipolar affective disorder.  
XX  
XX Homo sapiens.  
XX  
XX MO200163294-A2.  
XX  
XX 30-AUG-2001.  
XX  
XX 23-FEB-2001; 2001WO-GB00791.  
XX  
XX 24-FEB-2000; 2000GB-0004412.  
XX PR 08-DEC-2000; 2000GB-0030030.  
XX PR 12-DEC-2000; 2000US-0254830.  
XX  
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX  
XX Herath HMAc, Parekh RB, Rohlf C;  
XX  
XX WPI: 2001-582081/65.  
XX  
XX Preparation for diagnosing or treating bipolar affected disorder (BAD)  
XX or unipolar depression, or for screening for modulators, comprises a  
XX BAD-associated protein isoform -  
XX  
XX Claim 8; Page 31; 163pp; English.  
XX  
XX The invention relates to a preparation comprising an isolated bipolar  
XX affected disorder (BAD)-Associated Protein Isoform (DPIs). The DPI's are  
XX used to screen, diagnose or prognose of BAD or unipolar depression,  
XX determine the stage or severity of BAD or unipolar depression, identify a  
XX subject at risk of developing BAD or unipolar depression, or monitor the  
XX effect of therapy in a subject. They are also used to screen for or  
XX identify agents that interact with a DPI. These agents, antibodies  
XX against the DPIs, and nucleic acids encoding the DPIs are used to treat  
XX or prevent BAD or unipolar depression. Diseases that can be treated are  
XX attention deficient disorder, a schizoaffective disorder, a bipolar or a  
XX unipolar affective disorder. The DPIs are used in proteomics. The

CC proteomic approach of using DPIs for screening, diagnosis or prognosis of  
CC BAD or unipolar depression overcomes the problems of using gene  
CC expression analysis, such as not being able to obtain central nervous  
CC system (CNS) tissue from a living patient under normal circumstances.  
CC The present sequence is a DIP decreased in the CSF (cerebro-spinal  
CC fluid) of subjects having BAD.  
XX  
SO Sequence 9 AA:  
  
Query Match 51.2%; Score 22; DB 22; Length 9;  
Best Local Similarity 80.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 DFTLT 6  
| | | | |  
Db 2 dftls 6  
  
RESULT 12  
AAV40735  
ID AAV40735 standard; peptide: 7 AA.  
XX  
XX AAV40735;  
XX  
DT 01-DEC-1999 (first entry)  
XX  
XX S4 derivative #9, beta strand of scaffold protein structure.  
XX  
XX Scaffold protein; beta strand; beta sheet; stabilize antigen; vaccine;  
XX tumour; chemotherapeutic agent.  
XX  
XX Synthetic.  
XX  
XX EP947582-A1.  
XX  
XX 06-OCT-1999.  
XX  
XX 31-MAR-1998; 98BP-0870065.  
XX  
XX 31-MAR-1998; 98BP-0870065.  
XX  
XX (INNO-) INNOGENETICS NV.  
XX  
XX Desmet J, Hufon S, Hoogenboom H, Sablon E;  
XX  
XX WPI: 1999-542958/46.  
XX  
XX New scaffold protein, useful for stabilizing antigens used as vaccines  
XX -  
XX  
XX Disclosure: Page 6; 105pp; English.  
XX  
XX Sequences AAV40727-Y40748 are functionally equivalent derivatives of the  
XX S4 peptide (AAV40607) which forms part of a scaffold protein. S4 is a  
XX beta strand peptide which forms part of a beta sheet. Peptides  
XX (AAV40601-Y40609) together form a single-chain scaffold protein which  
XX contains at least 1 disulfide bond, contains less than 10% alpha helix  
XX and contains at least 6 beta-strands. The scaffold protein is constructed  
XX of beta strands S1-S6, and may also include beta strands A1-A3, or any  
XX functionally equivalent derivative of these sequences. The beta strands  
XX form two beta sheets S1/S4/S3 and S6/S5/S2 with each strand connected to  
XX the next by hydrogen bonds, which generate a beta sandwich architecture.  
XX If the additional beta strands A1-A3 are included in the structure the  
XX scaffold is constructed of two beta sheets, with the structures  
XX A1/S1/S4/S3 and S6/S5/S2/A2/A3. The beta strands are connected to each  
XX other via amino acid loops, where at least one of the loops binds to a  
XX receptor or antigen. The scaffold protein is used to stabilize antigens  
XX or whole proteins such as receptors, or their fragments. It may be used  
XX to bind two separate molecules. For example, one surface of the scaffold  
XX may be bound to a protein which binds to a tumour antigen. This will  
XX target the complex to tumour cells. Another surface may be bound to a  
XX cytotoxic molecule or an autoimmune antibody which may then kill the





PD 13-JUL-1995.  
 XX  
 PF 06-JAN-1995; 95MO-US00147.  
 XX  
 PR 07-JAN-1994; 94US-0178570.  
 XX  
 PA (BETH-) BETH ISRAEL HOSPITAL ASSOC.  
 XX  
 PI Cantley LC, Songyang Z;  
 XX  
 DR WPI: 1995-255036/33.  
 XX  
 PT Determn. of amino acid sequence of protein kinase phosphorylation  
 PT site - by phosphorylation of peptide library and sequencing  
 PT phospho:peptides) formed, also new substrates and their analogues  
 PT for modulating or detecting protein kinase  
 XX  
 PS Example 5: Page 32; 131pp: English.  
 XX  
 CC An oriented degenerate peptide library of the amino acid formula  
 CC AAR79661 was constructed to isolate the amino acid sequences at the  
 CC phosphorylation sites of a protein kinase eg. protein kinase A,  
 CC cyclin B/p33(cdc2), src family kinases, etc. Peptides which are  
 CC phosphorylated are isolated and their amino acid sequences are compared  
 CC to known substrate/inhibitor peptide sequences for that protein kinase.  
 CC The peptides AAR79662-73 represent phosphorylation sites for protein  
 CC kinase A. This peptide sequence is the phosphorylation site in glycogen  
 CC synthase site 2.  
 CC The isolated peptides can be used to screen cpds. for effects on the  
 CC protein kinase activity, generate antibodies to identify native kinase  
 CC substrates, or modulate a variety of cellular responses in which protein  
 CC kinases are involved eg. cell cycle control, immune response,  
 CC transcriptional activation or cell development.  
 XX  
 SQ Sequence 9 AA;

Query Match 48.88; Score 21; DB 16; Length 9;  
 Best Local Similarity 66.7%; Pred. No. 6.4e+05;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 4 TLTSS 9  
 ||:||  
 Db 3 t1svs 8

Search completed: July 15, 2002, 13:42:57  
 Job time: 207 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:41:15 ; Search time 12.84 Seconds  
(without alignments)  
17.121 Million cell updates/sec

Title: US-09-712-819A-7

Perfect score: 43

Sequence: 1 TDFLTRISS 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 60703

Minimum DB seq length: 0

Maximum DB seq length: 9

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/prodata/2/1aa/5A.COMB.pep:\*

2: /cgn2\_6/prodata/2/1aa/5B.COMB.pep:\*

3: /cgn2\_6/prodata/2/1aa/6A.COMB.pep:\*

4: /cgn2\_6/prodata/2/1aa/6B.COMB.pep:\*

5: /cgn2\_6/prodata/2/1aa/PCUTS.COMB.pep:\*

6: /cgn2\_6/prodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	53.5	8	2	US-08-656-177A-14
2	23	53.5	8	2	US-09-256-797-14
3	21	48.8	9	1	US-08-178-570-22
4	21	48.8	9	3	US-08-369-643-22
5	21	48.8	9	5	PCT-US95-00147-22
6	20	46.5	9	2	US-08-934-222-103
7	20	46.5	9	2	US-08-933-407-103
8	20	46.5	9	2	US-09-207-621-103
9	20	46.5	9	2	US-08-532-818-103
10	20	46.5	9	3	US-09-231-797-103
11	20	46.5	9	3	US-08-934-224-103
12	20	46.5	9	3	US-08-933-843-103
13	20	46.5	9	3	US-08-934-223-103
14	20	46.5	9	4	US-09-413-492-103
15	20	46.5	9	4	US-09-187-859-3749
16	19	44.2	6	4	US-09-140-084-4
17	19	44.2	7	3	US-08-916-43A-8
18	19	44.2	7	4	US-08-640-737-38
19	19	44.2	8	1	US-08-276-452A-81
20	19	44.2	8	2	US-08-798-744-81
21	19	44.2	8	4	US-09-187-859-1755
22	19	44.2	8	4	US-09-187-859-2977
23	19	44.2	8	5	PCT-US94-05150-11
24	19	44.2	9	2	US-08-529-190B-67
25	19	44.2	9	3	US-08-159-339A-275
26	19	44.2	9	4	US-08-772-282-21
27	19	44.2	9	4	US-09-187-859-2978

28	19	44.2	9	4	US-09-187-859-3935	Sequence 3935, Ap
29	19	44.2	9	5	PCT-US95-02121-3	Sequence 3, Appl1
30	18	41.9	7	4	US-09-187-859-637	Sequence 637, Ap
31	18	41.9	8	4	US-09-400-208B-18	Sequence 18, Appl
32	18	41.9	8	4	US-09-187-859-638	Sequence 638, Ap
33	18	41.9	8	4	US-09-187-859-640	Sequence 640, Ap
34	18	41.9	8	4	US-09-187-859-2507	Sequence 2507, Ap
35	18	41.9	8	6	5496924-22	Patent No. 5496924
36	18	41.9	9	1	US-08-245-853-9	Sequence 9, Appl1
37	18	41.9	9	2	US-08-573-675-9	Sequence 87, Appl
38	18	41.9	9	3	US-08-986-234-87	Sequence 910, Ap
39	18	41.9	9	3	US-08-159-339A-910	Sequence 911, Ap
40	18	41.9	9	4	US-08-159-339A-911	Sequence 641, Ap
41	18	41.9	9	4	US-09-187-859-641	Sequence 2508, Ap
42	18	41.9	9	4	US-09-187-859-2508	Sequence 3683, Ap
43	18	41.9	9	4	US-09-187-859-3683	Sequence 3835, Ap
44	18	41.9	9	4	US-09-187-859-3855	Patent No. 5496924
45	18	41.9	9	6	5496924-36	

## ALIGNMENTS

RESULT 1  
US-08-656-177A-14  
; Sequence 14, Application US/08656177A  
; Patent No. 5882851  
; GENERAL INFORMATION:  
; APPLICANT: Koch, Birgit M.  
; APPLICANT: Sibbesen, Ole  
; APPLICANT: Halkier, Barbara Ann  
; APPLICANT: Moller, Birger L.  
; TITLE OF INVENTION: Cytochrome P-450 Monooxygenases  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5882851artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: NC  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/656,177A  
; FILING DATE: 08-AUG-1996  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meligs, J. Timothy  
; REGISTRATION NUMBER: 38, 241  
; REFERENCE/DOCKET NUMBER: S-19783/A/PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYDROTHERICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: Internal  
; US-08-656-177A-14

Query Match 53.5%; Score 23; DB 2; Length 8;  
Best Local Similarity 80.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 DFTLT 6  
|1|:1  
Db 2 DFTVT 6

## RESULT 2

US-09-256-797-14

Sequence 14, Application US/09256797  
Patent No. 6133417  
GENERAL INFORMATION:  
APPLICANT: Koch, Birgit M.  
APPLICANT: Sidsen, Ole  
APPLICANT: Halter, Barbara Ann  
APPLICANT: Moller, Birger L.  
TITLE OF INVENTION: Cytochrome P-450 Monooxygenases  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6133417artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/256,797  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/656,177  
FILING DATE: 08-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: S-19783/A/PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
US-09-256-797-14

Query Match 53.5%; Score 23; DB 4; Length 8;  
Best Local Similarity 80.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 DFTLT 6  
|1|:1  
Db 2 DFTVT 6

## RESULT 3

US-08-178-570-22

Sequence 22, Application US/08178570  
Patent No. 5532167  
GENERAL INFORMATION:  
APPLICANT: Lewis C. Cantley  
APPLICANT: Zhou Song Yang  
TITLE OF INVENTION: Substrate Specificity of Protein Kinases  
NUMBER OF SEQUENCES: 77  
CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 STATE STREET, suite 510  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02109-1875

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/178,570  
FILING DATE: JANUARY 7, 1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: DeConti, Giulio A., Jr.  
REGISTRATION NUMBER: 31,503  
REFERENCE/DOCKET NUMBER: BBI-004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: Internal  
US-08-178-570-22

Query Match 48.8%; Score 21; DB 1; Length 9;  
Best Local Similarity 66.7%; Pred. No. 1.7e+05;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 TLTVSS 9  
|1|:1|  
Db 3 TLTVSS 8

RESULT 4  
US-08-369-643-22  
Sequence 22, Application US/08369643A  
Patent No. 6004757  
GENERAL INFORMATION:  
APPLICANT: Cantley, Lewis C.  
APPLICANT: Songyang, Zhou  
TITLE OF INVENTION: Substrate Specificity of Protein Kinases  
FILE REFERENCE: CNS-001CP  
CURRENT APPLICATION NUMBER: US/08/369,643A  
CURRENT FILING DATE: 1995-01-06  
EARLIER APPLICATION NUMBER: US 08/178,570  
EARLIER FILING DATE: 1994-01-07  
NUMBER OF SEQ ID NOS: 92  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 22  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:glycogen  
US-08-369-643-22

Query Match 48.8%; Score 21; DB 3; Length 9;  
Best Local Similarity 66.7%; Pred. No. 1.7e+05;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 TLTVSS 9  
|1|:1|  
Db 3 TLTVSS 8

```
RESULT 5
PCT-US95-00147-22
; Sequence 22, Application PC/TUS9500147
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Substrate Specificity of Protein Kinases
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, suite 510
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/00147
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/178,570
; FILING DATE: JANUARY 7, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: DeConti, Giulio A., Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBT-004CPPC
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: Internal
; PCT-US95-00147-22

Query Match 48.8%; Score 21; DB 5; Length 9;
Best Local Similarity 66.7%; Pred. No. 1.7e+05;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 TLTISS 9
|:|:|
Db 3 TLSVSS 8

RESULT 6
US-08-934-222-103
; Sequence 103, Application US/08934222
; Patent No. 5928896
; GENERAL INFORMATION:
; APPLICANT: EVANS, Herbert J.
; APPLICANT: KINI, R. Manjunatha
; TITLE OF INVENTION: Polypeptides That Include Conformation-
; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interaction
; TITLE OF INVENTION: Site
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: Suite 500, 3000 K Street NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20007
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934,222
FILING DATE: 19-SEPT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/532,818
FILING DATE: 03-MAY-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/143,364
FILING DATE: 29-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/051,741
FILING DATE: 23-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Isaacson, John P.
REGISTRATION NUMBER: 33,751
REFERENCE/DOCKET NUMBER: 040433/0148
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
TOPOLOGY: linear

US-08-934-222-103

Query Match 46.5%; Score 20; DB 2; Length 9;
Best Local Similarity 80.0%; Pred. No. 1.7e+05;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 TLTISS 8
|:|:|
Db 3 TLTISS 7

RESULT 7
US-08-933-402-103
; Sequence 103, Application US/08933402
; Patent No. 5948887
; GENERAL INFORMATION:
; APPLICANT: EVANS, Herbert J.
; APPLICANT: KINI, R. Manjunatha
; TITLE OF INVENTION: Polypeptides That Include Conformation-
; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interacti
; TITLE OF INVENTION: Site
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: Suite 500, 3000 K Street NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20007
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,402
; FILING DATE: 19-SEPT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/532,818
; FILING DATE: 03-MAY-1996
; APPLICATION NUMBER: U.S. 08/143,364
; FILING DATE: 29-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/051,741
; FILING DATE: 23-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Isaacson, John P.
```

REGISTRATION NUMBER: 33,751  
REFERENCE/DOCKET NUMBER: 040433/0148  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-933-402-103

Query Match 46.5%; Score 20; DB 2; Length 9;  
Best Local Similarity 80.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 TLTLS 8  
|||:|  
Db 3 TLTLS 7

RESULT 8  
US-09-207-621-103  
Sequence 103, Application US/09207621  
Patent No. 5952465  
GENERAL INFORMATION:  
APPLICANT: EVANS, Herbert J.  
TITLE OF INVENTION: Polypeptides That Include Conformation-  
CONSTRAINING GROUPS THAT INCLUDE FLANK A PROTEIN-PROTEIN INTERACTION  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: Suite 500, 3000 K Street NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20007  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/207,621  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/532,818  
FILING DATE: 03-MAY-1996  
APPLICATION NUMBER: PCT/US94/04294  
FILING DATE: 21-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/143,364  
FILING DATE: 29-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/051,741  
FILING DATE: 23-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Isaacson, John P.  
REGISTRATION NUMBER: 33,751  
REFERENCE/DOCKET NUMBER: 040433/0148  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-09-207-621-103

Query Match 46.5%; Score 20; DB 2; Length 9;  
Best Local Similarity 80.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 TLTLS 8  
|||:|

Db 3 TLTLS 7

RESULT 9  
US-08-532-818-103  
Sequence 103, Application US/08532818  
Patent No. 5965698  
GENERAL INFORMATION:  
APPLICANT: EVANS, Herbert J.  
TITLE OF INVENTION: Polypeptides That Include Conformation-  
CONSTRAINING GROUPS THAT INCLUDE FLANK A PROTEIN-PROTEIN INTERACTI  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: Suite 500, 3000 K Street NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20007  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/532,818  
FILING DATE: 03-MAY-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/04294  
FILING DATE: 21-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/143,364  
FILING DATE: 29-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/051,741  
FILING DATE: 23-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Isaacson, John P.  
REGISTRATION NUMBER: 33,751  
REFERENCE/DOCKET NUMBER: 040433/0148  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-532-818-103

Query Match 46.5%; Score 20; DB 2; Length 9;  
Best Local Similarity 80.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 TLTLS 8  
|||:|  
Db 3 TLTLS 7

RESULT 10  
US-09-231-797-103  
Sequence 103, Application US/09231797  
Patent No. 6084066  
GENERAL INFORMATION:  
APPLICANT: EVANS, Herbert J.  
TITLE OF INVENTION: Polypeptides That Include Conformation-  
CONSTRAINING GROUPS THAT INCLUDE FLANK A PROTEIN-PROTEIN INTERACTIO  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: Suite 500, 3000 K Street NW  
CITY: Washington

STATE: DC  
COUNTRY: USA  
ZIP: 20007  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/231,797  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/532,818  
FILING DATE: 03-MAY-1996  
APPLICATION NUMBER: PCT/US94/04294  
FILING DATE: 21-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/143,364  
FILING DATE: 29-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/051,741  
FILING DATE: 23-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Isaacson, John P.  
REGISTRATION NUMBER: 33,751  
REFERENCE/DOCKET NUMBER: 040433/0148  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-09-231-797-103

Query Match 46.5%; Score 20; DB 3; Length 9;  
Best Local Similarity 80.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 TLITS 8  
|||:|  
Db 3 TLITS 7

RESULT 11  
US-08-934-224-103  
Sequence 103, Application US/08934224  
Patent No. 6100044  
GENERAL INFORMATION:  
APPLICANT: EVANS, Herbert J.  
APPLICANT: KINI, R. Manjunatha  
TITLE OF INVENTION: Polypeptides That Include Conformation-  
TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interaction  
TITLE OF INVENTION: Site  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: Suite 500, 3000 K Street NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20007  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/934,224  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/532,818  
FILING DATE: 03-MAY-1996  
APPLICATION NUMBER: PCT/US94/04294

FILING DATE: 21-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/143,364  
FILING DATE: 29-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/051,741  
FILING DATE: 23-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Isaacson, John P.  
REGISTRATION NUMBER: 33,751  
REFERENCE/DOCKET NUMBER: 040433/0148  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-934-224-103

Query Match 46.5%; Score 20; DB 3; Length 9;  
Best Local Similarity 80.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 TLITS 8  
|||:|  
Db 3 TLITS 7

RESULT 12  
US-08-933-843-103  
Sequence 103, Application US/08933843  
Patent No. 6111069  
GENERAL INFORMATION:  
APPLICANT: EVANS, Herbert J.  
APPLICANT: KINI, R. Manjunatha  
TITLE OF INVENTION: Polypeptides That Include Conformation-  
TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interacti  
TITLE OF INVENTION: Site  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: Suite 500, 3000 K Street NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20007  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/933,843  
FILING DATE: 19-SEPT-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/532,818  
FILING DATE: 03-MAY-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/143,364  
FILING DATE: 29-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/051,741  
FILING DATE: 23-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Isaacson, John P.  
REGISTRATION NUMBER: 33,751  
REFERENCE/DOCKET NUMBER: 040433/0148  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-933-843-103

Query Match 46.5%; Score 20; DB 3; Length 9;  
Best Local Similarity 80.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 TLITS 8  
| | | : |  
DB 3 TLITS 7

RESULT 13  
US-08-934-223-103  
; Sequence 103, Application US/08934223  
; Patent No. 6147189  
; GENERAL INFORMATION:  
; APPLICANT: EVANS, Herbert J.  
; TITLE OF INVENTION: Polypeptides That Include Conformation-  
; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interaction  
; TITLE OF INVENTION: Site  
; NUMBER OF SEQUENCES: 153  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: Suite 500, 3000 K Street NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20007  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/934,223  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/532,818  
; FILING DATE: 03-MAY-1996  
; APPLICATION NUMBER: PCT/US94/04294  
; FILING DATE: 21-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: U.S. 08/143,364  
; FILING DATE: 29-OCT-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: U.S. 08/051,741  
; FILING DATE: 23-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Isaacson, John P.  
; REGISTRATION NUMBER: 33,751  
; REFERENCE/DOCKET NUMBER: 040433/0148  
; INFORMATION FOR SEQ ID NO: 103:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-08-934-223-103

Query Match 46.5%; Score 20; DB 4; Length 9;  
Best Local Similarity 80.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 TLITS 8  
| | | : |  
DB 3 TLITS 7

RESULT 14  
US-09-413-492-103  
; Sequence 103, Application US/09413492  
; Patent No. 6258550

GENERAL INFORMATION:  
; APPLICANT: EVANS, Herbert J.  
; APPLICANT: KINI, R. Manjunatha  
; TITLE OF INVENTION: Polypeptides That Include Conformation-  
; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interaction  
; TITLE OF INVENTION: Site  
; NUMBER OF SEQUENCES: 153  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: Suite 500, 3000 K Street NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20007  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/413,492  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/532,818  
; FILING DATE: 03-MAY-1996  
; APPLICATION NUMBER: PCT/US94/04294  
; FILING DATE: 21-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: U.S. 08/143,364  
; FILING DATE: 29-OCT-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: U.S. 08/051,741  
; FILING DATE: 23-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Isaacson, John P.  
; REGISTRATION NUMBER: 33,751  
; REFERENCE/DOCKET NUMBER: 040433/0148  
; INFORMATION FOR SEQ ID NO: 103:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-09-413-492-103

Query Match 46.5%; Score 20; DB 4; Length 9;  
Best Local Similarity 80.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 TLITS 8  
| | | : |  
DB 3 TLITS 7

RESULT 15  
US-09-187-859-3749  
; Sequence 3749, Application US/09187859A  
; Patent No. 6358920  
; GENERAL INFORMATION:  
; APPLICANT: Blaschuk, Orest W.  
; APPLICANT: Gour, Barbara J.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL  
; TITLE OF INVENTION: CAMBRIAN-MEDIATED FUNCTIONS  
; FILE REFERENCE: 100086,407C1  
; CURRENT APPLICATION NUMBER: US/09/187,859A  
; CURRENT FILING DATE: 1998-11-06  
; NUMBER OF SEQ ID NOS: 4052  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 3749  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:



OTHER INFORMATION: Representative cyclic modulating agent based on  
OTHER INFORMATION: cadherin-related neuronal receptor cell adhesion  
OTHER INFORMATION: recognition sequence  
us-09-187-859-3749

Query Match 46.5%; Score 20; DB 4; Length 9;  
Best Local Similarity 50.0%; Pred. No. 1.7e+05;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Qy 2 DFTLTISS 9  
:11: 11  
Db 1 EFTIDSS 8

Search completed: July 15, 2002, 13:43:16  
Job time: 121 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:45:26 ; Search time 14.1 Seconds  
(without alignments)  
74.963 Million cell updates/sec

Title: US-09-712-819A-8  
Perfect score: 58  
Sequence: 1 YGRKKRRQKK 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 1325

Minimum DB seq length: 0  
Maximum DB seq length: 11

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71:\*\n1: PIR1:\*\n2: PIR2:\*\n3: PIR3:\*\n4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	41.4	10	2 A42089	transcription fact
2	20	34.5	10	2 P00785	NADH dehydrogenase
3	18	31.0	11	2 I54193	Rhesus blood group
4	17	29.3	9	2 PC7076	spectrin alpha cna
5	17	29.3	10	2 A33143	pneumadin - rat
6	17	29.3	11	2 JQ0395	hypothetical prote
7	16	27.6	5	2 PQ0689	photosystem I 10.4
8	16	27.6	7	2 B33541	hypothetical prote
9	16	27.6	10	2 A60410	beta-neoendorphin
10	15	25.9	7	2 F33932	Ig mu chain D regl
11	15	25.9	9	2 S78762	ribosomal protein
12	15	25.9	10	2 S65387	cytochrome-c oxid
13	15	25.9	10	2 I52645	gene B-50 protein
14	15	25.9	11	2 B43669	hypothetical prote
15	14	24.1	6	2 I51434	H4 histone - Afri
16	14	24.1	9	2 C36730	leucosulfakinin-II
17	14	24.1	10	1 GMROL2	leucosulfakinin-II
18	14	24.1	10	2 S38305	lectin GNL2 alpha
19	14	24.1	10	2 B60656	leucosulfakinin II
20	14	24.1	10	2 S71948	matrix metalloprot
21	14	24.1	10	2 PH0923	T-cell receptor be
22	14	24.1	10	4 S14943	UAG3 leader peptid
23	14	24.1	11	1 GMROL	leucosulfakinin -
24	14	24.1	11	1 A40693	transgelin - sheep
25	14	24.1	11	2 A44755	20alpha-hydroxyste
26	14	24.1	11	2 A60656	perisulfakinin - A
27	14	24.1	11	2 PT0250	Ig heavy chain CRD
28	14	24.1	11	2 A33571	folistatin - bovi
29	13	22.4	5	2 B61445	Leu-enkephalin - b

30	13	22.4	5	2 A61445	Met-enkephalin - b
31	13	22.4	6	2 A60494	antineoplastic gly
32	13	22.4	7	2 A60139	fatty-acid synthas
33	13	22.4	7	2 A60224	Met-enkephalin-Arg
34	13	22.4	8	2 P00701	unidentified 6.5/3
35	13	22.4	8	2 PT0311	Ig heavy chain CRD
36	13	22.4	8	2 PT0323	Ig heavy chain CRD
37	13	22.4	8	2 PH1618	Ig H chain V-D-J r
38	13	22.4	9	2 E28854	transaldolase B -
39	13	22.4	9	2 A11497	Ig heavy chain CRD
40	13	22.4	9	2 PT0238	Ig heavy chain CRD
41	13	22.4	9	2 PT0285	Ig heavy chain CRD
42	13	22.4	9	2 PT0324	Ig heavy chain CRD
43	13	22.4	9	2 PH1591	Ig H chain V-D-J r
44	13	22.4	9	2 S36850	Ig heavy chain V r
45	13	22.4	9	2 G41946	T-cell receptor ga

## ALIGNMENTS

RESULT 1  
A42089  
Transcription factor I-POU protein, alternative splice form - fruit fly (Drosophila m  
C:Species: Drosophila melanogaster  
C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Mar-1998  
C:Accession: A42089  
R:Treacy, M.N.; Neilson, L.I.; Turner, E.E.; He, X.; Rosenfeld, M.G.  
Cell 68, 491-505, 1992  
A>Title: Twin of I-POU: a two amino acid difference in the I-POU homeodomain disting  
A:Reference number: A42089; MUID:92154665  
A:Accession: A42089  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-10 <TR>  
A:Cross-references: GB:S82271; NID:9245517; PID:9245518  
A>Note: sequence extracted from NCHI backbone (NCBIN:82267, NCBIN:82271, NCBIP:82273)  
C:Genetics:  
A:Gene: FlyBase:lpou  
A:Cross-references: FlyBase:FBgn0004418

Query Match 41.4%; Score 24; DB 2; Length 10;  
Best Local Similarity 57.1%; Pred. No. 6.5e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GRRKRRQ 8  
Db 1 GRRKRR 7

RESULT 2  
P00785  
NADH dehydrogenase (EC 1.6.99.3) 27K chain - fava bean mitochondrion (fragment)  
M:Alternate names: complex I 27K chain; NADH-ubiquinone reductase 27K chain  
C:Species: mitochondrion Vicia faba (fava bean)  
C>Date: 03-May-1994 #sequence\_revision 07-Oct-1994 #text\_change 17-Mar-1999  
R:Leclercq, S.; Boutry, M.  
Plant Physiol. 102, 435-443, 1993  
A>Title: Purification and preliminary characterization of mitochondrial complex I (NA  
A:Reference number: P00775; MUID:94151437  
A:Accession: P00785  
A:Molecule type: protein  
A:Residues: 1-10 <LEP>  
C:Comment: Complex I, mitochondrial NADH-ubiquinone reductase, is the first of the  
ranging from 5K to 75K.  
C:Comment: This enzyme catalyzes electron transfer from endogenous NADH to ubiquinone  
A:Genome: mitochondrion  
C:Keywords: electron transfer; mitochondrion; oxidoreductase

Query Match 34.5%; Score 20; DB 2; Length 10;  
 Best Local Similarity 50.0%; Pred. No. 2.8e+03;  
 Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 GRRKRR 7  
 ||||  
 DB 5 GRRKRR 10

RESULT 3  
 154193  
 Rhesus blood group CcDe protein - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 06-Sep-1996 #sequence\_revision 06-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: 154193  
 R:Cherif-zaher, B.; Le Van Kim, C.; Rouillac, C.; Reynal, V.; Carton, J.P.; Collin, Y.  
 Genomics 19, 68-74, 1994  
 A:Title: Organization of the gene (RRCDE) encoding the human blood group RhCcDe antigens  
 A:Reference number: 154193; MUID:94245182  
 A:Accession: 154193  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-11 <RES>  
 A:Cross-references: GB:S70456; NID:9546795; PIDN:AD014061.1; PID:94261761  
 C:Genetics:  
 A:Gene: GDB:RHCE  
 A:Cross-references: GDB:229957; OMIM:111700  
 A:Map position: 1p36.2-1p34

Query Match 31.0%; Score 18; DB 2; Length 11;  
 Best Local Similarity 57.1%; Pred. No. 6.2e+03;  
 Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 YGRKKRR 7  
 ||||  
 DB 5 YPRSVRR 11

RESULT 4  
 PC7076  
 spectrin alpha chain, non-erythroid - mouse (fragment)  
 N:Alternate names: fodrin alpha chain  
 C:Species: Mus musculus (house mouse)  
 C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
 C:Accession: PC7076  
 R:Tsuigita, A.; Kawakami, T.; Uchida, T.; Sakai, T.; Kamo, M.; Matsui, T.; Watanabe, Y.;  
 Electrophoresis 21, 1853-1871, 2000  
 A:Title: Proteome analysis of mouse brain: Two-dimensional electrophoresis profiles of  
 A:Reference number: PC7076  
 A:Accession: PC7076  
 A:Molecule type: protein  
 A:Residues: 1-9 <TSU>  
 A:Experimental source: strain C57BL/6Cr SLc, male; brain, striatum  
 C:Keywords: brain

Query Match 29.3%; Score 17; DB 2; Length 9;  
 Best Local Similarity 75.0%; Pred. No. 2.8e+03;  
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 RKKR 6  
 ||||  
 DB 5 RKKX 8

RESULT 5  
 A33143  
 pneumadin - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 16-Nov-1990 #sequence\_revision 16-Nov-1990 #text\_change 18-Aug-2000  
 C:Accession: A33143  
 R:Batra, V.K.; Mathur, M.; Mitr, S.A.; Kapoor, R.; Kumar, M.A.

Regul. Pept. 30, 77-87, 1990  
 A:Title: Pneumadin: a new lung peptide which triggers antidiuretic.  
 A:Reference number: A33143; MUID:9110910  
 A:Accession: A33143  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-10 <BAT>  
 C:Superfamily: unassigned animal peptides

Query Match 29.3%; Score 17; DB 2; Length 10;  
 Best Local Similarity 60.0%; Pred. No. 8.2e+03;  
 Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YGRKK 5  
 ||||  
 DB 1 YGEPK 5

RESULT 6  
 J00395  
 hypothetical protein (nodB 3' region) - Azorhizobium caulinodans  
 N:Alternate names: hypothetical 3.4k protein  
 C:Species: Azorhizobium caulinodans  
 A:Note: Host Sesbania rostrata  
 C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 03-Feb-1994  
 C:Accession: J00395  
 R:Goethals, K.; Gao, M.; Tomekpe, K.; Van Montagu, M.; Holsters, M.  
 Mol. Gen. Genet. 219, 289-298, 1989  
 A:Title: Common nodABC genes in Nod locus 1 of Azorhizobium caulinodans: nucleotide s  
 A:Reference number: J00395; MUID:90136519  
 A:Accession: J00395  
 A:Molecule type: DNA  
 A:Residues: 1-11 <GOE>  
 A:Cross-references: GB:U18897  
 A:Experimental source: strain ORS571

Query Match 29.3%; Score 17; DB 2; Length 11;  
 Best Local Similarity 66.7%; Pred. No. 8.8e+03;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 RROKKK 11  
 ||||  
 DB 2 RRVKK 7

RESULT 7  
 P00689  
 photosystem I 10.4k H1 chain - common tobacco (fragment)  
 C:Species: Nicotiana tabacum (common tobacco)  
 C:Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 17-Mar-1999  
 C:Accession: P00689  
 R:Obokata, J.; Mikami, K.; Hayashida, N.; Nakamura, M.; Sugitara, M.  
 Plant Physiol. 102, 1259-1267, 1993  
 A:Title: Molecular heterogeneity of photosystem I. psad, psae, psaf, psah and psal ar  
 A:Reference number: P00689; MUID:94105345  
 A:Accession: P00689  
 A:Molecule type: protein  
 A:Residues: 1-5 <OBO>  
 C:Keywords: chloroplast; photosynthesis; photosystem I; thylakoid

Query Match 27.6%; Score 16; DB 2; Length 5;  
 Best Local Similarity 75.0%; Pred. No. 2.8e+03;  
 Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 YGRK 4  
 ||||  
 DB 2 YGDK 5

RESULT 8

B33541  
 hypothetical protein (T1 5' region) - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 21-Feb-1990 #sequence\_revision 21-Feb-1990 #text\_change 30-Sep-1993  
 C:Accession: B33541  
 R:Klemenz, R.; Hoffmann, S.; Werenskiold, A.K.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 5708-5712, 1989  
 A:Title: Serum- and oncoprotein-mediated induction of a gene with sequence similarity to  
 A:Reference number: A33541; MUID:89345536  
 A:Accession: B33541  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-7 <KLE>  
 A:Cross-references: GB:M24843

Query Match 27.6%; Score 16; DB 2; Length 7;  
 Best Local Similarity 50.0%; Pred. No. 2.8e+05;  
 Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 5 KRQKK 10  
 |||  
 Db 2 RRRSAK 7

RESULT 9  
 A60410  
 beta-neoendorphin / dynorphin precursor - guinea pig  
 N:Alternate names: alpha-neoendorphin; proenkephalin B precursor  
 C:Species: Cavia porcellus (guinea pig)  
 C>Date: 03-Feb-1993 #sequence\_revision 03-Feb-1993 #text\_change 21-Jan-2000  
 C:Accession: A60410  
 R:Murphy, R.; Turner, C.A.  
 Peptides 11, 65-68, 1990  
 A:Title: Isolation and microsequence analysis of guinea pig alpha-neo-endorphin.  
 A:Reference number: A60410; MUID:90259864  
 A:Accession: A60410  
 A:Molecule type: protein  
 A:Residues: 1-10 <MR>  
 C:Superfamily: proenkephalin  
 C:Keywords: neuropeptide; opioid peptide

Query Match 27.6%; Score 16; DB 2; Length 10;  
 Best Local Similarity 40.0%; Pred. No. 1.2e+04;  
 Matches 4; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 YGKKRKQKK 10  
 |||  
 Db 1 YGFLRKRPK 10

RESULT 10  
 E33932  
 Ig mu chain D region (E7) - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 09-Mar-1990 #sequence\_revision 09-Mar-1990 #text\_change 16-Aug-1996  
 C:Accession: E33932  
 R:Baccala, R.; Vo Quang, T.; Gilbert, M.; Ternynck, T.; Avrameas, S.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 4624-4628, 1989  
 A:Title: Two murine natural polyreactive autoantibodies are encoded by nonmutated germ-1  
 A:Reference number: A33932; MUID:89282823  
 A:Accession: E33932  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-7 <BAC>  
 A:Cross-references: GB:M27106  
 C:Keywords: immunoglobulin

Query Match 25.9%; Score 15; DB 2; Length 7;  
 Best Local Similarity 66.7%; Pred. No. 2.8e+05;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 YGR 3  
 |||  
 Db 4 YGR 6

RESULT 11  
 S78762  
 ribosomal protein MRP-S12, mitochondrial - bovine (fragment)  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: S78762  
 R:Graack, H.R.  
 Submitted to the Protein Sequence Database, July 1999  
 A:Reference number: S78760  
 A:Accession: S78762  
 A:Molecule type: protein  
 A:Residues: 1-9 <GRA>  
 C:Keywords: mitochondrial  
 F:1-9/Product: ribosomal protein MRP-S12 (fragment) #status experimental <MAT>

Query Match 25.9%; Score 15; DB 2; Length 9;  
 Best Local Similarity 42.9%; Pred. No. 2.8e+05;  
 Matches 3; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 RKKRKQK 9  
 |||  
 Db 2 RKKRPR 8

RESULT 12  
 S65387  
 cytochrome-c oxidase (EC 1.9.3.1) chain VII b, cardiac - rat (fragment)  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 16-Jul-1999  
 C:Accession: S65387; S65386  
 R:Schaeffer, H.; Noack, H.; Halanek, W.; Brandt, U.; von Jagow, G.  
 Eur. J. Biochem. 230, 235-241, 1995  
 A:Title: Cytochrome-c oxidase in developing rat heart. Enzymic properties and amino-t  
 A:Reference number: S65372; MUID:95324529  
 A:Accession: S65387  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-10 <SCH>  
 A:Accession: S65386  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-10 <SC2>  
 C:Keywords: cardiac muscle; heart; oxidoreductase

Query Match 25.9%; Score 15; DB 2; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+04;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 QKK 10  
 |||  
 Db 3 QKK 5

RESULT 13  
 I52645  
 gene B-50 protein - rat (fragment)  
 C:Species: Rattus sp. (rat)  
 C>Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 17-Mar-1999  
 C:Accession: I52645  
 R:Pegen, B.J.; Melander, H.B.; Rensen-de Leeuw, M.G.; Scholtman, P.; Gispén, W.H.; Sc  
 Brain Res. Mol. Brain Res. 23, 221-234, 1994  
 A:Title: Identification of two promoter regions in the rat B-50/GAP-43 gene.  
 A:Reference number: I52645; MUID:94335554  
 A:Accession: I52645  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Residues: 1-10 <RES>  
 A:Cross-references: GB:S71492; NID:9560728  
 C:Genetics:  
 A:Gene: B-50

Query Match 25.9%; Score 15; DB 2; Length 10;  
 Best Local Similarity 60.0%; Pred. No. 1.7e+04;  
 Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 6 RRQK 10  
 |||  
 Db 6 RRQK 10

RESULT 14  
 B43669  
 hypothetical protein (rhdA 5' region) - *Synechococcus* sp. (fragment)  
 C:Species: *Synechococcus* sp.  
 C>Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 30-Sep-1993  
 C:Accession: B43669  
 R:Laudenbach, D.E.; Ehrhardt, D.; Green, L.; Grossman, A.  
 J. Bacteriol. 173, 2751-2760, 1991  
 A:Title: Isolation and characterization of a sulfur-regulated gene encoding a periplasmic  
 A:Reference number: A43669; MUID:91210163  
 A:Accession: B43669  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-11 <LANO>  
 A:Cross-references: GB:M65244

Query Match 25.9%; Score 15; DB 2; Length 11;  
 Best Local Similarity 75.0%; Pred. No. 1.8e+04;  
 Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 KRRQ 8  
 |||  
 Db 7 KNRQ 10

RESULT 15  
 I51434  
 H4 histone - African clawed frog (fragment)  
 C:Species: *Xenopus laevis* (African clawed frog)  
 C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: I51434  
 R:Woodland, H.R.; Warrington, J.R.; Ballantyne, J.E.M.; Turner, P.C.  
 Nucleic Acids Res. 12, 4939-4958, 1984  
 A:Title: Are there major developmentally regulated H4 gene classes in *Xenopus*?  
 A:Reference number: I51391; MUID:B4247348  
 A:Accession: I51434  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-6 <MOO>  
 A:Cross-references: GB:K02304; NID:9214227; PIDN:AAA49738.1; PID:9555517

Query Match 24.1%; Score 14; DB 2; Length 6;  
 Best Local Similarity 75.0%; Pred. No. 2.8e+05;  
 Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GRKK 5  
 |||  
 Db 3 GRGK 6

Search completed: July 15, 2002, 13:47:29  
 Job time: 123 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:47:11 ; Search time 10.15 Seconds

(without alignments)  
41.962 Million cell updates/sec

Title: US-09-712-819A-8

Sequence: 1 YGRRRRRQKK 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 36719550 residues

Total number of hits satisfying chosen parameters: 410

Minimum DB seq length: 0  
Maximum DB seq length: 11

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	29.3	8	1	UH09_RAT
2	17	29.3	10	1	PNEU_RAT
3	17	29.3	11	1	UXR2_YEAST
4	15	25.9	10	1	COXM_RAT
5	14	24.1	9	1	HUTU_KLEAE
6	14	24.1	9	1	NSK1_SARBO
7	14	24.1	10	1	LSK2_LEUMA
8	14	24.1	10	1	SPI_HAIRO
9	14	24.1	11	1	LSK1_LEUMA
10	14	24.1	11	1	LSKP_PERAM
11	13	22.4	8	1	UN06_PINPS
12	13	22.4	8	1	AL16_CARMA
13	13	22.4	9	1	FIRB_PAPPA
14	13	22.4	9	1	TAL3_PICHA
15	13	22.4	9	1	TKC1_CALVO
16	13	22.4	9	1	TKL1_LOCM1
17	13	22.4	10	1	CU30_LOCM1
18	13	22.4	10	1	GONI_ALUM1
19	13	22.4	10	1	GONI_ONCKE
20	13	22.4	10	1	TKL2_LOCM1
21	13	22.4	10	1	TKL3_LOCM1
22	13	22.4	10	1	TKN1_SCYCA
23	13	22.4	10	1	TKNB_RANCA
24	13	22.4	10	1	TKS1_AEDAE
25	13	22.4	10	1	TKS2_AEDAE
26	13	22.4	11	1	CA31_LITCI
27	13	22.4	11	1	CA32_LITCI
28	13	22.4	11	1	TKN1_UPEIN
29	13	22.4	11	1	TKN1_UPERU
30	13	22.4	11	1	TKN2_UPERU
31	13	22.4	11	1	TKNA_GADMO
32	13	22.4	11	1	TKNA_HORSE
33	13	22.4	11	1	TKNA_RANCA

34	13	22.4	11	1	TKNA_RANRI	P29207 rana ridibu
35	13	22.4	11	1	TKN_PHYFU	P08615 physalaemus
36	12	20.7	9	1	B44K_PORGI	P81886 porphyromon
37	12	20.7	8	1	FARP_CALSI	P38495 callinectes
38	12	20.7	10	1	RL16_ACHLA	P29221 acholeplasm
39	12	20.7	11	1	CEP1_ACHFU	P23790 achelina fu
40	12	20.7	11	1	COR2_PERAM	P11496 periplaneta
41	11	19.0	3	1	GRWM_HUMAN	P01157 homo sapien
42	11	19.0	5	1	UF01_MOUSE	P38639 mus musculu
43	11	19.0	7	1	FARS_HIRME	P42564 hirudo medi
44	11	19.0	9	1	FIRB_ERYPA	P19346 erythrocebu
45	11	19.0	9	1	FIRB_MACFU	P19345 macaca fusc

## ALIGNMENTS

RESULT	ID	Query Match	Score	DB 1	Length	DB 1
1	UH09_RAT	Best Local Similarity	29.3%	Score 17	8	DB 1
AC	P56575	Matches 3; Conservative	2	Mismatches 0	Gaps 0	Indels 0
DT	15-DEC-1998 (rel. 37, last sequence update)					
DT	15-DEC-1998 (rel. 37, last sequence update)					
DT	15-DEC-1998 (rel. 37, last sequence update)					
DE	Unknown protein from 2D-page of heart tissue (Spot P9) (Fragment).					
OS	Rattus norvegicus (Rat).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
NCBI_TaxID-10116;						
RM	[1]					
RP	SEQUENCE:					
RC	STRAIN=WISTAR; TISSUE=Heart;					
RA	Li X.-P., Pleissner K.-P., Scheler C., Regitz-Zagrosek V., Salikov J.,					
RA	Jungblut P.R.;					
RL	Submitted (SEP-1998) to the SWISS-PROT data bank.					
CC	-i- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN					
CC	PROTEIN IS: 8.9, ITS MW IS: 42 KDa.					
FT	NON_TER					
EQ	SEQUENCE 8 AA: 1029 MW; 9R0775A6C4140B06 CRC64;					

Query Match	Score	DB 1	Length	DB 1
Best Local Similarity	29.3%	Score 17	8	DB 1
Matches 3; Conservative	2	Mismatches 0	Gaps 0	Indels 0
DB	1	QERRQ 5		
QY	4	KRRQ 8		
DB	1	QERRQ 5		
RESULT	2			
ID	PNEU_RAT	STANDARD:	PRT:	10 AA.
AC	P21996:			
DT	01-AUG-1991 (rel. 19, created)			
DT	01-AUG-1991 (rel. 19, last sequence update)			
DT	16-OCT-2001 (rel. 40, last annotation update)			
DE	Pneumadin (PNM).			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID-10116;			
RM	[1]			
RP	SEQUENCE, AND SYNTHESIS.			
RC	TISSUE=Lung;			
RC	MEDLINE=91110910; PubMed=2274681;			
RT	Bacteria V.K., Machur M., Mir S.A., Kapoor R., Kumar M.A.;			
RT	"Pneumadin: a new lung peptide which triggers antidiuresis."			
RL	Regul. Pept. 30:77-87(1990).			
CC	-i- FUNCTION: THIS ANTIIDIURETIC PEPTIDE TRIGGERS THE RELEASE OF ADH.			
CC	PIR: A33143; A33143.			
DR	PIR: A33143; A33143.			
KW	AMIDATION.			
FT	MOD_RES 10 10 AMIDATION.			

5Q SEQUENCE 10 AA; 1048 MW; 6A1D0DDAA723276B CRC64;

Query Match 29.3%; Score 17; DB 1; Length 10;  
Best Local Similarity 60.0%; Pred. No. 1.6e+03;  
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YGRK 5  
DB 1 YCEPK 5

RESULT 3  
UXB2\_YEAST

ID UXB2\_YEAST STANDARD; PRT; 11 AA.

AC P99013;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Unknown protein from 2D-page (Spot 2D-000K2F) (Fragment).  
OS Saccharomyces cerevisiae (Baker's yeast).  
CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
ON NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE.  
RC STRAIN-X2180-1A;  
RA Sanchez J.-C., Golaz O., Schaller D., Morch F., Frutiger S.,  
RH Hughes G.J., Appel R.D., Deshusses J., Hochstrasser D.F.;  
RL Submitted (MUG-1995) to the SWISS-PROT data bank.  
CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN  
DR PROTEIN IS: 6.20, ITS MW IS: 9.2 kDa.  
CC SWISS-2DPAGE: P99013; YEAST.  
FT NON\_TER 11 11  
SQ SEQUENCE 11 AA; 1328 MW; EC38021C0DCBA42DA CRC64;

Query Match 29.3%; Score 17; DB 1; Length 11;  
Best Local Similarity 75.0%; Pred. No. 1.8e+03;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 YGRK 4  
DB 6 YARK 9

RESULT 4  
COXM\_RAT

ID COXM\_RAT STANDARD; PRT; 10 AA.

AC P80431;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Cytochrome c oxidase polypeptide VIIb, mitochondrial (EC 1.9.3.1)  
DE (Fragment).  
GN COX7B.  
OS Rattus norvegicus (Rat).  
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
ON NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE.  
RC STRAIN-MISTAR; TISSUE=Liver;  
RX MEDLINE=95324529; PubMed=7601105;  
RA Schaeffer H., Noack H., Halanek W., Brandt U., von Jagow G.;  
RT "Cytochrome c oxidase in developing rat heart. Enzymic properties and  
amino-terminal sequences suggest identity of the fetal heart and the  
adult liver isoform."  
RL Eur. J. Biochem. 230:235-241(1995).  
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE  
CHAINS OF CYTOCHROME C OXIDASE, THE TERMINAL OXIDASE IN  
MITOCHONDRIAL ELECTRON TRANSPORT.  
CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) -> 4 ferricyclochrome

CC c + 2 H(2)O.  
KM Oxidoreductase; Mitochondrion.  
FT NON\_TER 10 10

5Q SEQUENCE 10 AA; 1210 MW; CFC70EB771A33326 CRC64;

Query Match 25.9%; Score 15; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 3.4e+03;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 OKK 10  
DB 3 OKK 5

RESULT 5  
HUTU\_KLEAE

ID HUTU\_KLEAE STANDARD; PRT; 9 AA.

AC P12381;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Urocanate hydratase (EC 4.2.1.49) (Urocanase) (imidazolonepropionate  
DE hydrolase) (Fragment).  
GN HUTU.  
OS Klebsiella aerogenes.  
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
CC Klebsiella.  
ON NCBI\_TaxID=28451;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88198018; PubMed=2834335;  
RA Nieukoop A.J., Baldauf S.A., Hudepeth M.E.S., Bender R.A.;  
RT "bidirectional promoter in the hut(P) region of the histidine  
RT utilization (hut) operons from Klebsiella aerogenes."  
RL J. Bacteriol. 170:2240-2246(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90368611; PubMed=2203754;  
RA Schwacha A., Bender R.A.;  
RT "Nucleotide sequence of the gene encoding the repressor for the  
RT histidine utilization genes of Klebsiella aerogenes."  
RL J. Bacteriol. 172:5477-5481(1990).  
CC -1- CATALYTIC ACTIVITY: 4,5-dihydro-4-oxo-5-imidazolepropionate -  
urocanate + H(2)O.  
CC -1- COFACTOR: 1 TIGHTLY BOUND NAD(+) PER CATALYTIC SUBUNIT (BY  
SIMILARITY).  
CC -1- PATHWAY: SECOND STEP IN HISTIDINE DEGRADATION.  
CC -1- SIMILARITY: BELONGS TO THE UROCANASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M19665; AAA25078.1; -  
CC EMBL: M34604; AAA25076.1; -  
DR PIR: C36730; C36730.  
DR InterPro: IPR000193; Urocanase.  
DR PROSITE: PS01233; UROCANASE; PARTIAL.  
KW Histidine metabolism; Lyase; NAD.  
FT NON\_TER 9 9  
SQ SEQUENCE 9 AA; 1140 MW; 970FC41B5325A6C5 CRC64;

Query Match 24.1%; Score 14; DB 1; Length 9;  
Best Local Similarity 50.0%; Pred. No. 1e+05;  
Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 RKKRQ 8



```

Db      3 QSKYRQ 8

RESULT 6
ID      NSK1_SARBU      STANDARD:      PRT:      9 AA.
AC      P41492;
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DE      Neosulfakinin-I (NEB-SK-I).
OS      Sarcophaga bullata (Grey flesh fly) (Neobellieria bullata).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Oestroidea; Sarcophagidae; Sarcophaga.
OX      NCBI_TaxID=7385;
RM      [1]
RP      SEQUENCE.
RC      TISSUE=Head;
RX      MEDLINE=93083101; PubMed=1360367;
RA      Fonagy A., Schoofs L., Proost P., van Damme J., de Loof A.;
RT      *Isolation and primary structure of two sulfakinin-like peptides from
RT      the fleshfly, Neobellieria bullata.*;
RL      Comp. Biochem. Physiol. 103C:135-142(1992).
CC      -1- FUNCTION: MYOTROPIC PEPTIDE.
CC      -1- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
DR      InterPro: IPR001651; Gastrin.
DR      PROSITE: PS00259; GASTRIN; 1.
KW      Neuropeptide; Amidation; Sulfation.
FT      MOD_RES      4      4      Sulfation (POTENTIAL).
FT      MOD_RES      9      9      AMIDATION (POTENTIAL).
SO      SEQUENCE      9 AA; 1187 MW; 8B0A0691E86B5A0A CRC64;

Query Match      24.1%; Score 14; DB 1; Length 9;
Best Local Similarity 40.0%; Pred. No. 1e+05; 2; Indels 0; Gaps 0;
Matches 2; Conservative 1; Mismatches 2;

QY      1 YGRKK 5
Db      4 YGHMR 8

RESULT 7
ID      LSK2_LEUMA      STANDARD:      PRT:      10 AA.
AC      P09039;
DT      01-NOV-1988 (Rel. 09, Created)
DT      01-FEB-1994 (Rel. 28, Last sequence update)
DT      01-NOV-1995 (Rel. 32, Last annotation update)
DE      Leucosulfakinin-II (LSK-II).
OS      Leucophaea maderae (Madefira cockroach), and
OS      Periplaneta americana (American cockroach).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC      Blaberoidae; Blaberidae; Leucophaea.
OX      NCBI_TaxID=6988, 6978;
RM      [1]
RP      SEQUENCE.
RC      SPECIES=L.maderae;
RX      MEDLINE=87048769; PubMed=3778455;
RA      Nachman R.J., Holman G.M., Cook B.J., Haddon W.F., Ling N.;
RT      *Leucosulfakinin-II, a blocked sulfated insect neuropeptide with
RT      homology to cholecystokinin and gastrin.*;
RL      Biochem. Biophys. Res. Commun. 140:357-364(1986).
RM      [2]
RP      SEQUENCE.
RC      SPECIES=P.americana; TISSUE=Corpora cardiaca;
RX      MEDLINE=90137190; PubMed=2615921;
RA      Veenstra J.A.;
RT      *Isolation and structure of two gastrin/CCK-like neuropeptides from
RT      the American cockroach homologous to the leucosulfakinins.*;

```

```

RL      Neuropeptides 14:145-149(1989).
CC      -1- FUNCTION: CHANGE THE FREQUENCY AND AMPLITUDE OF CONTRACTIONS OF
CC      THE COCKROACH HINDGUT. STIMULATES MUSCLE CONTRACTION OF HINDGUT.
CC      -1- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
DR      PIR; A26335; GMR02.
DR      PIR; B60656; B60656.
DR      InterPro: IPR001651; Gastrin.
DR      PROSITE: PS00259; GASTRIN; 1.
KW      Hormone; Amidation; Sulfation.
FT      MOD_RES      1      1      PYROLIDONE CARBOXYLIC ACID.
FT      MOD_RES      5      5      SUFFATION (IN L.MADERAE, BUT NOT IN
FT      MOD_RES      1      1      P.AMERICANA).
SO      SEQUENCE      10 AA; 1255 MW; 9B4F5391E86B5A0A CRC64;

Query Match      24.1%; Score 14; DB 1; Length 10;
Best Local Similarity 40.0%; Pred. No. 5e+03;
Matches 2; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 YGRKK 5
Db      5 YGHMR 9

RESULT 8
ID      SPL_HALRO      STANDARD:      PRT:      10 AA.
AC      Q10987;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      15-DEC-1998 (Rel. 37, Last annotation update)
DE      Serine proteinase inhibitor (Fragment).
OS      Halocynthia roretzi (Sea squirt).
OS      Halocynthia metazoa; Chordata; Urochordata; Ascidiacea;
OC      Eukaryota; Metazoa; Chordata; Pyrosida; Halocynthia.
OX      NCBI_TaxID=7729;
RM      [1]
RP      SEQUENCE.
RC      TISSUE=Hemolymph;
RX      MEDLINE=96321313; PubMed=8759295;
RA      Shishikura F., Abe T., Ohtake S.-I., Tanaka K.;
RT      *Purification and characterization of a 58,000-Da proteinase
RT      inhibitor from the hemolymph of a solitary ascidian, Halocynthia
RT      roretzi.*;
RL      Comp. Biochem. Physiol. 114B:1-9(1996).
CC      -1- FUNCTION: STRONGLY INHIBITS TRYPSIN AND PLASMA ENZYME(S) ACTIVITY.
CC      -1- SUBUNIT: MONOMER.
CC      -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR      InterPro: IPR000215; Serpin.
DR      PROSITE: PS00284; SERPIN; PARTIAL.
KW      Serpin; Serine protease inhibitor; Glycoprotein; Plasma.
FT      NON_TER      10      10
FT      SEQUENCE      10 AA; 1104 MW; 4225C73B1B187AA3 CRC64;

Query Match      24.1%; Score 14; DB 1; Length 10;
Best Local Similarity 42.9%; Pred. No. 5e+03; 2; Indels 0; Gaps 0;
Matches 3; Conservative 2; Mismatches 2;

QY      4 KKRQKK 10
Db      2 KKDGEK 8

RESULT 9
ID      LSK1_LEUMA      STANDARD:      PRT:      11 AA.
AC      P04428;
DT      13-AUG-1987 (Rel. 05, Created)
DT      13-AUG-1987 (Rel. 05, Last sequence update)
DT      01-NOV-1995 (Rel. 32, Last annotation update)
DE      Leucosulfakinin-I (LSK-I).

```

OS Leucophaea maderae (Mediterranean cockroach).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;  
 OC Blaberidae; Blaberidae; Leucophaea.  
 OX NCBI\_TaxID=6988;  
 RN (1)  
 RP SEQUENCE.  
 RA MEDLINE=86315858; PubMed=3749893;  
 RA Nachman R.J., Holman G.M., Haddon W.F., Ling N.;  
 RT "Leucosulafakinin, a sulfated insect neuropeptide with homology to  
 RT gastrin and cholecystokinin.";  
 RL Science 234:71-73(1986).  
 CC -1- FUNCTION: CHANGE THE FREQUENCY AND AMPLITUDE OF CONTRACTIONS OF  
 CC THE COCKROACH HINDGUT. INHIBITS MUSCLE CONTRACTION OF HINDGUT.  
 CC -1- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.  
 DR PIR: A01622; GMR0L.  
 DR InterPro: IPR001651; Gastrin.  
 DR PROSITE: PS00259; GASTRIN.1.  
 KW Hormone; Amidation; Sulfation.  
 FT MOD\_RES 6 6 SULFATION.  
 FT MOD\_RES 11 11 AMIDATION.  
 SQ SEQUENCE 11 AA: 1459 MW: 784E0680E86B5AAB CRC64:

Query Match 24.1%; Score 14; DB 1; Length 11;  
 Best Local Similarity 40.0%; Pred. No. 5.4e+03;  
 Matches 2; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YGRKK 5  
 11  
 DB 6 YGHR 10

RESULT 10  
 LSKP\_PERAM STANDARD; PRT; 11 AA.  
 AC P36885;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Periplaneta (Pee-Sk-1).  
 DE Periplaneta americana (American cockroach).  
 OS Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;  
 OC Blattidae; Blattidae; Periplaneta.  
 OX NCBI\_TaxID=6978;  
 RN (1)  
 RP SEQUENCE.  
 RC TISSUE=Corpora cardiaca;  
 RX MEDLINE=90137190; PubMed=2615921;  
 RA Veenstra J.A.;  
 RT "Isolation and structure of two gastrin/CKK-like neuropeptides from  
 RT the American cockroach homologous to the leucosulafakinins.";  
 RL Neuropeptides 14:145-149(1989).  
 CC -1- FUNCTION: STIMULATES HINDGUT CONTRACTIONS.  
 CC -1- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.  
 DR PIR: A06056; A06056.  
 DR InterPro: IPR001651; Gastrin.  
 DR PROSITE: PS00259; GASTRIN.1.  
 KW Hormone; Amidation; Sulfation.  
 FT MOD\_RES 6 6 SULFATION.  
 FT MOD\_RES 11 11 AMIDATION.  
 SQ SEQUENCE 11 AA: 1445 MW: 8B4E0680E86B5AAB CRC64:

Query Match 24.1%; Score 14; DB 1; Length 11;  
 Best Local Similarity 40.0%; Pred. No. 5.4e+03;  
 Matches 2; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YGRKK 5  
 11  
 DB 6 YGHR 10

RESULT 11  
 UN06\_PINPS STANDARD; PRT; 7 AA.  
 AC P81675;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Unknown protein from 2D page of needles (N141) (Fragment).  
 OS Pinus pinaster (Maritime pine).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferales; Pinaceae; Pinus.  
 OX NCBI\_TaxID=71647;  
 RN (1)  
 RP SEQUENCE.  
 RC TISSUE=Needle;  
 RX MEDLINE=99274088; PubMed=10344291;  
 RA Costa P., Pionneau C., Bauw G., Dubos C., Bahman N., Kremer A.,  
 RA Frigerio J.-M., Plomion C.;  
 RT "Separation and characterization of needle and xylem maritime pine  
 RT proteins.";  
 RL Electrophoresis 20:1098-1108(1999).  
 CC -1- MISCELLANEOUS: ON THE 2D-GEL, THE DETERMINED PI OF THIS UNKNOWN  
 CC PROTEIN IS: 6.6, ITS MW IS: 25 kDa.  
 FT NON\_TER 1 1  
 FT NON\_TER 7 7  
 SQ SEQUENCE 7 AA: 823 MW: 69D76724486B5740 CRC64:

Query Match 22.4%; Score 13; DB 1; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YG 2  
 11  
 DB 2 YG 3

RESULT 12  
 AL16\_CARMA STANDARD; PRT; 8 AA.  
 AC P81819;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Carcinus (Common shore crab) (Green crab).  
 OS Carcinus maenas (Common shore crab) (Green crab).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
 OC Eubrachyura; Portunoidae; Portunidae; Carcinus.  
 OX NCBI\_TaxID=6759;  
 RN (1)  
 RP SEQUENCE.  
 RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
 RX MEDLINE=98121193; PubMed=9461295;  
 RA Dove H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,  
 RA Thorpe A.;  
 RT "Isolation and identification of multiple neuropeptides of the  
 RT allatostatins superfamily in the shore crab Carcinus maenas.";  
 RL Eur. J. Biochem. 250:727-734(1997).  
 CC -1- FUNCTION: MAY ACT AS A NEUROTENSIN-LIKE OR NEUROMODULATOR.  
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 KW Neuropeptide; Amidation; Multigene family.  
 FT MOD\_RES 8 8 AMIDATION.  
 SQ SEQUENCE 8 AA: 813 MW: 7C286B45AB476878 CRC64:

Query Match 22.4%; Score 13; DB 1; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YG 2  
 11

```

Db      6 YG 7

RESULT 13
FIBR_PAPHA STANDARD: PRT; 9 AA.
ID FIBR_PAPHA
AC P19343;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Fibrinogen beta chain [Contains: Fibrinopeptide B] (Fragment).
GN FGB.
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Papio.
OX NCBI_TaxID=9557;
RN [1]
RP SEQUENCE.
RX MEDLINE=84161822; PubMed=6423621;
RA Nakamura S., Takemura O., Takahashi K.;
RT "Fibrinopeptides A and B of baboons (Papio anubis, Papio hamadryas,
RT and Theropithecus gelada): their amino acid sequences and
RT evolutionary rates and a molecular phylogeny for the baboons.";
RL J. Biochem. 94:1973-1978(1983).
CC -1- FUNCTION: FIBRINOGEN HAS A DOUBLE FUNCTION: YIELDING MONOMERS THAT
CC POLYMERIZE INTO FIBRIN AND ACTING AS A COPROCTOR IN PLATELET
CC AGGREGATION.
CC -1- SUBUNIT: HEXAMER CONTAINING 2 SETS OF 3 NONIDENTICAL CHAINS
CC (ALPHA, BETA AND GAMMA), LINKED TO EACH OTHER BY DISULFIDE BONDS.
CC -1- MISCELLANEOUS: CONVERSION OF FIBRINOGEN TO FIBRIN IS TRIGGERED BY
CC THROMBIN, WHICH CLEAVES FIBRINOPEPTIDES A AND B FROM ALPHA & BETA
CC CHAINS, AND THUS EXPOSES THE N-TERMINAL POLYMERIZATION SITES
CC RESPONSIBLE FOR THE FORMATION OF THE SOFT CLOT.
DR InterPro: IPR002181; Fibrinogen.C.
DR Prosite: PS00514; FIBRIN.AC_C_DOMAIN; PARTIAL.
KM BLOOD COAGULATION; Plasma.
FT PEPTIDE 1 9 FIBRINOPEPTIDE B.
FT NON_TER 9 9
SQ SEQUENCE 9 AA; 1057 MW; DDFE71E9C7287B06 CRC64;

Query Match 22.4%; Score 13; DB 1; Length 9;
Best Local Similarity 66.7%; Pred. No. 1e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YG 3
Db 7 HGR 9

RESULT 14
TAL3_PICUA STANDARD: PRT; 9 AA.
ID TAL3_PICUA
AC P17441;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transaldolase III (EC 2.2.1.2) (Fragment).
OS Pichia jadinii (Yeast) (Candida utilis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4903;
RN [1]
RP SEQUENCE.
RX MEDLINE=75145197; PubMed=1092268;
RA Teolae O., Sun S.C.;
RT "Isolation of a peptide containing a histidyl-L-cysteinyl sequence
RT from the active center of transaldolase.";
RL Arch. Biochem. Biophys. 167:525-533(1975).
CC -1- FUNCTION: TRANSALDOLASE IS IMPORTANT FOR THE BALANCE OF
CC METABOLITES IN THE PENTOSE-PHOSPHATE PATHWAY.

```

```

CC -1- CATALYTIC ACTIVITY: Sedoheptulose 7-phosphate + D-glyceraldehyde
CC 3-phosphate -> D-erythrose 4-phosphate + D-fructose 6-phosphate.
CC -1- PATHWAY: Pentose phosphate pathway; nonoxidative part.
CC -1- SIMILARITY: BELONGS TO THE TRANSALDOLASE FAMILY. SUBFAMILY 1.
DR PIR: A11497; A11497.
DR InterPro: IPR001585; Transaldolase.
DR PROSITE: PS00958; TRANSALDOLASE_2; PARTIAL.
DR PROSITE: PS01054; TRANSALDOLASE_1; PARTIAL.
KM Transferase; Pentose shunt.
FT NON_TER 1 1
FT NON_TER 9 9
SQ SEQUENCE 9 AA; 1033 MW; 325A31A4EB1E058 CRC64;

Query Match 22.4%; Score 13; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YG 2
Db 1 YG 2

RESULT 15
TKCL_CALVO STANDARD: PRT; 9 AA.
ID TKCL_CALVO
AC P41517;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Callitachykinin I.
OS Calliphora vomitoria (blue blowfly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Oestrolidae; Calliphoridae; Calliphora.
OX NCBI_TaxID=27454;
RN [1]
RP SEQUENCE. AND SYNTHESIS.
RX MEDLINE=95075727; PubMed=7984492;
RA Lundquist C.T., Clottens F.L., Holman G.M., Nichols R., Nachman R.J.,
RA Naessel D.R.;
RT "Callitachykinin I and II, two novel myotropic peptides isolated from
RT the blowfly, Calliphora vomitoria, that have resemblances to
RT tachykinins.";
RL Peptides 15:761-768(1994).
CC -1- FUNCTION: MYOACTIVE PEPTIDE.
CC -1- SIMILARITY: SOME SIMILARITY TO TACHYKININS.
KM Tachykinin; Neuropeptide; Amidation.
FT MOD_RES 9 9
FT SEQUENCE 9 AA; 981 MW; 2417C86B59C0C1B7 CRC64;

Query Match 22.4%; Score 13; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YG 2
Db 6 YG 7

Search completed: July 15, 2002, 13:50:14
Job time: 183 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:47:31 ; Search time 23.84 Seconds  
(Without alignments)  
79.821 Million cell updates/sec

Title: US-09-712-819A-8  
Perfect score: 58  
Sequence: 1 YGRRKRRQKK 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 1418

Minimum DB seq length: 0  
Maximum DB seq length: 11

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19:\*

- 1: sp.archaea:\*
- 2: sp.bacteria:\*
- 3: sp.fungi:\*
- 4: sp.human:\*
- 5: sp.invertebrate:\*
- 6: sp.mammal:\*
- 7: sp.mhc:\*
- 8: sp.organelle:\*
- 9: sp.phage:\*
- 10: sp.plant:\*
- 11: sp.podent:\*
- 12: sp.virus:\*
- 13: sp.vertibrate:\*
- 14: sp.unclassified:\*
- 15: sp.virus:\*
- 16: sp.bacteriap:\*
- 17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	36.2	9	2 Q93E20	Q93E20 streptococ
2	20	34.5	7	8 P92421	P92421 psathyrosta
3	20	34.5	7	8 P92385	P92385 hordeum mar
4	20	34.5	7	8 P92372	P92372 haynaldia v
5	20	34.5	7	8 P92403	P92403 lophopyrum
6	20	34.5	7	8 P92425	P92425 pseudoroegn
7	20	34.5	7	8 P92487	P92487 henrardia p
8	20	34.5	7	8 P92427	P92427 peridictyon
9	20	34.5	7	8 P92390	P92390 heteranthel
10	20	34.5	7	8 P92226	P92226 critlopsis
11	20	34.5	7	8 P92214	P92214 amblyopyrum
12	20	34.5	7	8 P92430	P92430 aegilops ta
13	20	34.5	7	8 P92421	P92421 bromus iner
14	20	34.5	7	8 P92442	P92442 taeniatheiru
15	20	34.5	7	8 P92381	P92381 hordeum bra
16	20	34.5	7	8 P92393	P92393 hordeum vul

17	20	34.5	7	8 P92218	P92218 australopyr
18	20	34.5	7	8 P92440	P92440 thlinopyrum
19	20	34.5	7	8 P92210	P92210 agropyron c
20	20	34.5	8	12 O84271	O84271 human papil
21	19	32.8	9	6 Q9T777	Q9T777 bos taurus
22	22	32.8	11	2 Q9E023	Q9E023 escherichia
23	18	31.0	8	12 Q84273	Q84273 human papil
24	18	31.0	10	11 Q9QVE5	Q9QVE5 mus sp. pro
25	18	31.0	10	15 Q86324	Q86324 rous sarcom
26	18	31.0	10	15 Q86325	Q86325 rous sarcom
27	18	31.0	10	15 Q86326	Q86326 rous sarcom
28	18	31.0	11	2 Q47345	Q47345 escherichia
29	17	29.3	7	12 Q9T1R0	Q9T1R0 human adeno
30	17	29.3	7	12 Q9Y1Q9	Q9Y1Q9 human adeno
31	17	29.3	7	12 Q9YVE3	Q9YVE3 human adeno
32	17	29.3	9	2 Q45852	Q45852 clostridium
33	17	29.3	9	10 Q940K4	Q940K4 arabidopsis
34	17	29.3	10	4 Q9UCU6	Q9UCU6 homo sapien
35	17	29.3	11	4 Q9Y3G2	Q9Y3G2 homo sapien
36	16	27.6	10	2 P83160	P83160 anabaena sp
37	16	27.6	10	2 Q51812	Q51812 escherichia
38	16	27.6	10	5 Q9V7K6	Q9V7K6 drosophila
39	16	27.6	11	5 Q9V8K1	Q9V8K1 mus musc
40	15	25.9	11	11 Q99M81	Q99M81 mus musc
41	15	25.9	9	11 Q99M83	Q99M83 mus musc
42	14	24.1	8	6 Q9R7U8	Q9R7U8 pseudomonas
43	14	24.1	8	6 Q9BF93	Q9BF93 megaptera n
44	14	24.1	8	6 Q9BF92	Q9BF92 tursiops tr
45	14	24.1	9	10 Q9AXH8	Q9AXH8 mesembryant
			12	Q82622	Q82622 avian infec

## ALIGNMENTS

RESULT 1

ID	Q93E20	PRELIMINARY:	PRT:	9 AA.
AC	Q93E20;			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	SCPB (FRAGMENT).			
GN	SCPB.			
OS	Streptococcus agalactiae.			
OC	Bacterial, Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
CC	Streptococcus.			
OX	NCBI_TaxID=1311;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Q90R;			
RA	MEDLINE=21424698; PubMed=11532154;			
FX	Franken C., Haase G., Brandt C., Weber-Heynemann J., Martin S.,			
RA	Lammier C., Podbielski A., Luitcken R., Spellerberg B.;			
RT	"Horizontal gene transfer and host specificity of beta-haemolytic			
RT	streptococci: the role of a putative composite transposon containing			
RT	scpb and lmb."			
RL	Mol. Microbiol. 41:925-935(2001).			
DR	EMBL: AF327852; MAF10713.1; ..			
FT	NON_TER			
SQ	SEQUENCE 9 AA; 1146 MW; 543721AB1326C403 CRC64;			

Query Match 36.2%; Score 21; DB 2; Length 9;  
Best Local Similarity 37.5%; Pred. No. 5.6e+05;  
Matches 3; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 KRRQKK 11  
1:1:1:1:1  
DB 2 KROKTKK 9

RESULT 2  
P92421

ID P92421 PRELIMINARY; PRT: 7 AA.  
 AC P92421; P92419;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last annotation update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11.  
 OS Psathyrostachys fragilis.  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Psathyrostachys.  
 OX NCBI\_TaxID=37729;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H4372, AND H917; TISSUE=LEAVES;  
 RA Petersen G., Seberg O.;  
 DR EMBL: Z77753; CAB01337.1; -;  
 DR EMBL: Z77752; CAB01334.1; -;  
 KW Ribosomal protein; Chloroplast.  
 FT NON\_TER 1  
 SQ SEQUENCE 7 AA: 894 MW; 673414033277700 CRC64;

Query Match 34.5%; Score 20; DB 8; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KRRR 7  
 ||||  
 DB 3 KRRR 6

RESULT 3  
 P92385 PRELIMINARY; PRT: 7 AA.  
 ID P92385; P92383;  
 AC P92385; P92383;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last annotation update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11.  
 OS Hordeum maritimum (seaside barley).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=4519;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H299, AND H801; TISSUE=LEAVES;  
 RA Petersen G., Seberg O.;  
 DR EMBL: Z77763; CAB01367.1; -;  
 DR EMBL: Z77762; CAB01364.1; -;  
 KW Ribosomal protein; Chloroplast.  
 FT NON\_TER 1  
 SQ SEQUENCE 7 AA: 894 MW; 673414033277700 CRC64;

Query Match 34.5%; Score 20; DB 8; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KRRR 7  
 ||||  
 DB 3 KRRR 6

RESULT 4  
 P92372 PRELIMINARY; PRT: 7 AA.  
 ID P92372

AC P92372;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11.  
 OS Haynaldia villosa.  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Palicouidae; Andropogoneae; Haynaldia.  
 OX NCBI\_TaxID=40247;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H5561; TISSUE=LEAF;  
 RC MEDLINE=97271648; PubMed=9126564;  
 RA Petersen G., Seberg O.;  
 RT "Phylogenetic analysis of the Triticeae (Poaceae) based on rpoA  
 sequence data."  
 RL Mol. Phylogenet. Evol. 7:217-230(1997).  
 DR EMBL: Z77741; CAB01301.1; -;  
 KW Chloroplast.  
 FT NON\_TER 1  
 SQ SEQUENCE 7 AA: 894 MW; 673414033277700 CRC64;

Query Match 34.5%; Score 20; DB 8; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KRRR 7  
 ||||  
 DB 3 KRRR 6

RESULT 5  
 P92403 PRELIMINARY; PRT: 7 AA.  
 ID P92403;  
 AC P92403;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11.  
 OS Lophopyrum elongatum (Tall wheatgrass) (Argopyrum elongatum).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Lophopyrum.  
 OX NCBI\_TaxID=4588;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H6692; TISSUE=LEAF;  
 RC MEDLINE=97271648; PubMed=9126564;  
 RA Petersen G., Seberg O.;  
 RT "Phylogenetic analysis of the Triticeae (Poaceae) based on rpoA  
 sequence data."  
 RL Mol. Phylogenet. Evol. 7:217-230(1997).  
 DR EMBL: Z77743; CAB01307.1; -;  
 KW Chloroplast.  
 FT NON\_TER 1  
 SQ SEQUENCE 7 AA: 894 MW; 673414033277700 CRC64;

Query Match 34.5%; Score 20; DB 8; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KRRR 7  
 ||||  
 DB 3 KRRR 6

RESULT 6  
 ID P92425 PRELIMINARY; PRT; 7 AA.  
 AC P92425;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11.  
 OS Pseudoeugenia spicata.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Pseudoeugenia.  
 OC NCBI\_TaxID=4604;  
 RX STRAIN=H9082; TISSUE=LEAF;  
 RC MEDLINE=97271648; PubMed=9126564;  
 RA Petersen G., Seberg O.;  
 RT "Phylogenetic analysis of the Triticeae (Poaceae) based on rpoA  
 sequence data."  
 RL MOL. Phylogenet. Evol. 7:217-230(1997).  
 DR EMBL: Z77744; CAB01310.1; -.  
 KW Chloroplast.  
 FT NON\_TER  
 SQ SEQUENCE 7 AA; 894 MW; 673414033277700 CRC64;

Query Match 34.5%; Score 20; DB 8; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KKRR 7  
 |||||  
 DB 3 KKRR 6

RESULT 7  
 ID P92387 PRELIMINARY; PRT; 7 AA.  
 AC P92387;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11.  
 OS Heterardia persica.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Hendaria.  
 OC NCBI\_TaxID=37678;  
 RX STRAIN=H5536; TISSUE=LEAF;  
 RC MEDLINE=97271648; PubMed=9126564;  
 RA Petersen G., Seberg O.;  
 RT "Phylogenetic analysis of the Triticeae (Poaceae) based on rpoA  
 sequence data."  
 RL MOL. Phylogenet. Evol. 7:217-220(1997).  
 DR EMBL: Z77748; CAB01322.1; -.  
 KW Chloroplast.  
 FT NON\_TER  
 SQ SEQUENCE 7 AA; 894 MW; 673414033277700 CRC64;

Query Match 34.5%; Score 20; DB 8; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KKRR 7  
 |||||

DB 3 KKRR 6  
 RESULT 8  
 ID P92427 PRELIMINARY; PRT; 7 AA.  
 AC P92427;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11.  
 OS Peridictyon sanctum.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Peridictyon.  
 OC NCBI\_TaxID=37683;  
 RX STRAIN=H5575; TISSUE=LEAF;  
 RC MEDLINE=97271648; PubMed=9126564;  
 RA Petersen G., Seberg O.;  
 RT "Phylogenetic analysis of the Triticeae (Poaceae) based on rpoA  
 sequence data."  
 RL MOL. Phylogenet. Evol. 7:217-230(1997).  
 DR EMBL: Z77749; CAB01325.1; -.  
 KW Chloroplast.  
 FT NON\_TER  
 SQ SEQUENCE 7 AA; 894 MW; 673414033277700 CRC64;

Query Match 34.5%; Score 20; DB 8; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KKRR 7  
 |||||  
 DB 3 KKRR 6

RESULT 9  
 ID P92390 PRELIMINARY; PRT; 7 AA.  
 AC P92390;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11.  
 OS Heteranthellium pilliferum.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Heteranthellium.  
 OC NCBI\_TaxID=37679;  
 RX STRAIN=H5557; TISSUE=LEAF;  
 RC MEDLINE=97271648; PubMed=9126564;  
 RA Petersen G., Seberg O.;  
 RT "Phylogenetic analysis of the Triticeae (Poaceae) based on rpoA  
 sequence data."  
 RL MOL. Phylogenet. Evol. 7:217-230(1997).  
 DR EMBL: Z77750; CAB01328.1; -.  
 KW Chloroplast.  
 FT NON\_TER  
 SQ SEQUENCE 7 AA; 894 MW; 673414033277700 CRC64;

Query Match 34.5%; Score 20; DB 8; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KRRR 7  
|||||  
DB 3 KRRR 6

RESULT 10  
P92226

PRELIMINARY: PRT: 7 AA.

AC P92226;  
DT 01-MAY-1997 (TREMblrel. 03, Created)  
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)  
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
GN RIBOSOMAL PROTEIN 11 (FRAGMENT).  
OS RPS11.  
OC Crithopsis delileana.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Trilicaceae; Crithopsis.  
OX NCBI\_TaxID=37674;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H5558; TISSUE=LEAF;  
RX MEDLINE=97271648; PubMed=9126564;  
RA Petersen G., Seberg O.;  
RT "Phylogenetic analysis of the Trilicaceae (Poaceae) based on rpoA  
sequence data.";  
RL Mol. Phylogenet. Evol. 7:217-230(1997).  
DR EMBL: Z77751; CAB01331.1; -.  
KW Chloroplast.  
FT NON\_TER  
SQ SEQUENCE 7 AA: 894 MW: 673414033277700 CRC64;

Query Match 34.5%; Score 20; DB 8; Length 7;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KRRR 7  
|||||  
DB 3 KRRR 6

RESULT 11  
P92214

PRELIMINARY: PRT: 7 AA.

AC P92214;  
DT 01-MAY-1997 (TREMblrel. 03, Created)  
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)  
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
GN RIBOSOMAL PROTEIN 11 (FRAGMENT).  
OS RPS11.  
OC Amblyopyrum multicum.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Trilicaceae; Amblyopyrum.  
OX NCBI\_TaxID=4595;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H5572; TISSUE=LEAF;  
RX MEDLINE=97271648; PubMed=9126564;  
RA Petersen G., Seberg O.;  
RT "Phylogenetic analysis of the Trilicaceae (Poaceae) based on rpoA  
sequence data.";  
RL Mol. Phylogenet. Evol. 7:217-230(1997).  
DR EMBL: Z77756; CAB01346.1; -.  
KW Chloroplast.  
FT NON\_TER  
SQ SEQUENCE 7 AA: 894 MW: 673414033277700 CRC64;

Query Match 34.5%; Score 20; DB 8; Length 7;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KRRR 7  
|||||  
DB 3 KRRR 6

RESULT 12  
P92430

PRELIMINARY: PRT: 7 AA.

AC P92430;  
DT 01-MAY-1997 (TREMblrel. 03, Created)  
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)  
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
GN RIBOSOMAL PROTEIN 11 (FRAGMENT).  
OS RPS11.  
OC Aegilops tauschii (Aegilops squarrosa).  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Trilicaceae; Aegilops.  
OX NCBI\_TaxID=37682;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H6668; TISSUE=LEAF;  
RX MEDLINE=97271648; PubMed=9126564;  
RA Petersen G., Seberg O.;  
RT "Phylogenetic analysis of the Trilicaceae (Poaceae) based on rpoA  
sequence data.";  
RL Mol. Phylogenet. Evol. 7:217-230(1997).  
DR EMBL: Z77758; CAB01352.1; -.  
KW Chloroplast.  
FT NON\_TER  
SQ SEQUENCE 7 AA: 894 MW: 673414033277700 CRC64;

Query Match 34.5%; Score 20; DB 8; Length 7;  
Best Local Similarity 100.0%; Pred. No. 5.6e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KRRR 7  
|||||  
DB 3 KRRR 6

RESULT 13  
P92221

PRELIMINARY: PRT: 7 AA.

AC P92221;  
DT 01-MAY-1997 (TREMblrel. 03, Created)  
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)  
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
GN RIBOSOMAL PROTEIN 11 (FRAGMENT).  
OS RPS11.  
OC Bromus inermis (Smooth brome grass).  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Bromaeae; Bromus.  
OX NCBI\_TaxID=13571;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-OSA414; TISSUE=LEAF;  
RX MEDLINE=97271648; PubMed=9126564;  
RA Petersen G., Seberg O.;  
RT "Phylogenetic analysis of the Trilicaceae (Poaceae) based on rpoA  
sequence data.";  
RL Mol. Phylogenet. Evol. 7:217-230(1997).  
DR EMBL: Z77759; CAB01355.1; -.  
KW Chloroplast.  
FT NON\_TER  
SQ SEQUENCE 7 AA: 894 MW: 673414033277700 CRC64;



SO SEQUENCE 7 AA; 894 MW; 673414033277700 CRC64;

## Query Match

Best Local Similarity 34.5%; Score 20; DB 8; Length 7;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KRRR 7  
 ||||  
 DB 3 KRRR 6

## RESULT 14

P92442 PRELIMINARY; PRT; 7 AA.

AC P92442;  
 DT 01-MAY-1997 (TREMBLERL. 03, Created)  
 DT 01-MAY-1997 (TREMBLERL. 03, Last sequence update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11.  
 OS Taeniatherum caput-medusae (Medusaehead).  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Taeniatherum.  
 OX NCBI\_TaxID=37687;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H10254; TISSUE-LEAF;  
 RA MEDLINE=97271648; PubMed=9126564;  
 RA Petersen G., Seberg O.;  
 RT "Phylogenetic analysis of the Triticeae (Poaceae) based on rpoA  
 sequence data.";  
 RL Mol. Phylogenet. Evol. 7:217-230(1997).  
 DR EMBL: 277760; CAB01358.1; -.  
 KW Chloroplast.  
 FT NON\_TER 1 1  
 SO SEQUENCE 7 AA; 894 MW; 673414033277700 CRC64;

## Query Match

Best Local Similarity 34.5%; Score 20; DB 8; Length 7;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KRRR 7  
 ||||  
 DB 3 KRRR 6

## RESULT 15

P92381 PRELIMINARY; PRT; 7 AA.

AC P92381;  
 DT 01-MAY-1997 (TREMBLERL. 03, Created)  
 DT 01-MAY-1997 (TREMBLERL. 03, Last sequence update)  
 DE RIBOSOMAL PROTEIN 11 (FRAGMENT).  
 GN RPS11.  
 OS Hordeum brachyantherum.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=52712;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LEAF;  
 RA MEDLINE=97271648; PubMed=9126564;  
 RA Petersen G., Seberg O.;  
 RT "Phylogenetic analysis of the Triticeae (Poaceae) based on rpoA  
 sequence data.";  
 RL Mol. Phylogenet. Evol. 7:217-230(1997).

DR EMBL: 277761; CAB01361.1; -.  
 KW Chloroplast.

FT NON\_TER 1 1  
 SO SEQUENCE 7 AA; 894 MW; 673414033277700 CRC64;

Query Match  
 Best Local Similarity 34.5%; Score 20; DB 8; Length 7;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KRRR 7  
 ||||  
 DB 3 KRRR 6

Search completed: July 15, 2002, 13:50:45  
 Job time: 194 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: Jul 15, 2002, 13:43:41 ; Search time 29.02 Seconds  
(Without alignments)  
42.102 Million cell updates/sec

Title: US-09-712-819a-8  
Perfect score: 58  
Sequence: 1 YGRRKRRQKK 11

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 147623

Minimum DB seq length: 0  
Maximum DB seq length: 11

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802: \*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of residues predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	89.7	11	22	AAU09906 Interleukin 17 (h1
2	49	84.5	11	19	AAW50263 HIV-1 tat protein.
3	49	84.5	11	20	AAV25075 TAT transduction d
4	49	84.5	11	20	AAV05415 TAT peptide. Unid
5	49	84.5	11	21	AAW03932 TAT protein trans
6	49	84.5	11	21	AAW03961 Minimal eleven ami
7	49	84.5	11	21	AAW27088 Beta-actinin deriv
8	49	84.5	11	21	AAW35698 Peptide associated
9	49	84.5	11	21	AAW29413 HIV TAT transducti
10	49	84.5	11	21	AAW09907 HIV tat protein tr
11	49	84.5	11	21	AAW93542 Amino acid sequenc

12	49	84.5	11	21	AAW71015 Human Immunodefici
13	49	84.5	11	22	AAU09932 HIV-1 tat protein
14	49	84.5	11	22	AAU09812 Protein transducti
15	49	84.5	11	22	AAE13064 Human Immunodefici
16	49	84.5	11	22	AAE12891 HIV tat protein fr
17	49	84.5	11	22	AAW56973 HIV-1 tat protein
18	49	84.5	11	22	AAW50221 Human Immunodefici
19	49	84.5	11	22	AAE12605 Membrane transport
20	49	84.5	11	22	AAE12204 Internalising pept
21	49	84.5	11	22	AAW60006 HIV TAT protein tr
22	49	84.5	11	22	AAW82757 HIV-1 tat protein
23	49	84.5	11	22	AAW85847 Human Immunodefici
24	49	84.5	11	22	AAW04300 Human Immunodefici
25	49	84.5	11	22	AAW05268 HIV TAT peptide fr
26	49	84.5	11	22	AAW89683 Human Immunodefici
27	49	84.5	11	22	AAW03418 Protein transducti
28	49	84.5	11	22	AAW03730 HIV tat protein tr
29	49	84.5	11	22	AAE02973 Protein transducti
30	49	84.5	11	22	AAW68376 Human G2 checkpoint
31	49	84.5	11	22	AAW67673 Transduction domai
32	49	84.5	11	22	AAW70458 HIV-1 TAT protein
33	49	84.5	11	22	AAW73305 HIV tat protein tr
34	49	84.5	11	22	AAW69170 HIV TAT protein tr
35	49	84.5	11	22	AAW89548 HIV TAT protein tr
36	49	84.5	11	22	AAW70481 NTR3 derived pepti
37	49	84.5	11	22	AAW71756 HIV TAT protein tr
38	49	84.5	11	22	AAW83627 HIV-1 Tat fragment
39	49	84.5	11	22	AAW83627 Tat derived peptid
40	49	77.6	10	21	AAW78335 HIV-1 Tat peptide
41	42	72.4	10	21	AAW78335 N-terminal S-benzyl
42	42	72.4	10	21	AAW85052 C-unn amino termin
43	42	72.4	10	22	AAW86312 HIV-1 TAT peptide.
44	42	72.4	10	22	AAW83193
45	42	72.4	10	22	AAW83193

#### ALIGNMENTS

RESULT 1	
AAU09906	standard; Protein: 11 AA.
ID	AAU09906
AC	AAU09906;
XX	
XX	14-FEB-2002 (first entry)
DT	
DE	Interleukin 17 (h1-17) receptor like protein associated peptide #1.
XX	
XX	Interleukin 17, h1-17 receptor like protein; immunomodulatory;
KW	anti-inflammatory; anti-diabetic; immunosuppressive; anti-microbial;
KW	hepatic; anabolic; anorectic; anti-alzheimer's; anti-parkinsonian;
KW	anti-convulsant; anti-asthmatic; dermatological; renal; osteopathic;
KW	vascular; cytostatic; anti-leukemic; anti-infertility; ophthalmologic;
KW	hepatitis; anorexia; cachexia; neuronal dysfunction; lung disease;
KW	bone disease; vascular disorder; eye disorder; cancer; human.
XX	
OS	Synthetic.
XX	
PN	W0200168859-A2.
XX	
PD	20-SEP-2001.
XX	
XX	15-MAR-2001; 2001MO-US06678.
PF	
XX	16-MAR-2000; 2000US-189816P.
PR	28-NOV-2000; 2000US-0724460.
XX	
PA	(AMGE-) AMGEN INC.
XX	
PI	Jing S;
XX	
DR	WPI; 2001-611392/70.

```

XX Nucleic acids encoding interleukin 17 receptor like polypeptides.
PT useful for preventing, diagnosing and treating, e.g. leukemia, asthma,
PT diabetes, psoriasis and glaucoma -
XX
XX Disclousure; Page 149; 158pp; English.
PS
XX The invention describes novel nucleic acids encoding interleukin (IL) 17
CC receptor like polypeptides useful as vaccines and in gene therapy. These
CC have immunomodulatory, anti-inflammatory, anti-diabetic, anti-microbial,
CC immunosuppressive, hepatic, anabolic, anorectic, anti-alzheimer's, renal,
CC anti-parkinsonian, anti-convasant, anti-asthmatic, dermatological,
CC osteopathic, vascular, cytostatic, anti-leukemic, anti-infectivity and
CC ophthalmological activities. The IL-17 receptor like nucleic acids and
CC proteins may be used to prevent and treat diseases associated with
CC inappropriate IL-17 receptor like polypeptide (IL17rip) expression. These
CC include, for example immune disorders (e.g. inflammation, diabetes and
CC transplant rejection), infections (e.g. hepatitis and septicemia),
CC weight disorders (e.g. anorexia, cachexia and obesity), neuronal
CC dysfunction (e.g. Alzheimer's disease, Parkinson's disease and epilepsy),
CC lung disorders (e.g. cystic fibrosis, asthma and emphysema), skin disease
CC (e.g. eczema and psoriasis), kidney disease (e.g. glomerulonephritis),
CC bone diseases (e.g. osteoporosis and hypercalcaemia), vascular disorders
CC (e.g. stroke and atherosclerosis, cancers (e.g. leukemia, myeloma and
CC breast cancer), reproductive disorders (e.g. infertility and
CC miscarriage), eye disorders (e.g. glaucoma and retinal neuropathy). The
CC DNA and its complements may also used as diagnostic probes to detect and
CC quantitate the presence of similar nucleic acids in samples and identify
CC patients needing restorative therapy. The IL17rip may also be used as
CC antigens in the production of antibodies against the proteins and in
CC assays to identify modulators of expression and activity. The
CC anti-IL17rip antibodies and antagonists may also be used to down regulate
CC expression and activity.
CC Note: This artificial peptide sequence is given in the sequence listing
CC but is not described in the specification.
XX
SO Sequence 11 AA:

Query Match 89.7%; Score 52; DB 22; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.042;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 YGRRKRRORKK 11
   |||||
DB 1 ygrtkrrgrkk 11

RESULT 2
AAW50263
ID AAW50263 standard; Protein; 11 AA.
AC AAW50263;
XX
XX 20-JUL-1998 (first entry)
DT
XX
XX HIV-1 tat protein.
DE
XX
XX Mouse; BH3 interacting domain death agonist; BID; BCL-2 family;
KM apoptosis; regulation; cell death; inflammation; cancer; arthritis;
KM autoimmune disease; viral infection; lymphoproliferative.
XX
XX Human immunodeficiency virus type 1.
OS
XX
XX MO980980-A1.
PN
XX
XX 12-MAR-1998.
PD
XX
XX 09-SEP-1997; 97WO-US15872.
PF
XX
XX 09-SEP-1996; 96US-0706741.
PR
XX
XX (UNITV ) UNITV WASHINGTON.
PA

```

```

XX Korsmeyer SJ;
PI
XX WPI; 1998-193546/17.
DR
XX
XX BH3 interacting domain death agonist polypeptide - used for treating
PT decreased apoptotic conditions resulting from inflammation etc.
PT
XX
XX Example 8; Page 85; 118pp; English.
PS
XX
XX The present sequence represents the HIV-1 tat protein which is used in
CC an example of the present invention which describes a BH3 interacting
CC domain death agonist (BID) truncated protein. The BID protein, the
CC DNA encoding it or antisense sequences can be used for preventing or
CC treating a decreased apoptotic state of a cell. The decreased apoptotic
CC state that is treated results from a disease such as cancer, viral
CC infections, lymphoproliferative conditions, arthritis, inflammation and
CC autoimmune diseases. Antibodies against the BID protein can be used for
CC detecting a BID polypeptide in a cell or population of cell. The nucleic
CC acid sequence and the BID protein can also be used for treating
CC immunodeficiency disease (including AIDS), senescence, neurodegenerative
CC disease, ischaemic and reperfusion cell death, infertility and
CC wound-healing. Primers derived from the nucleic acid encoding the BID
CC protein can be used for detecting/quantitating the protein and for
CC detecting alterations in the nucleic acid encoding the BID protein.
XX
SO Sequence 11 AA:

Query Match 84.5%; Score 49; DB 19; Length 11;
Best Local Similarity 72.7%; Pred. No. 0.12;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 YGRRKRRORKK 11
   |||||
DB 1 ygrtkrrgrrr 11

RESULT 3
AAV25075
ID AAV25075 standard; peptide; 11 AA.
XX
XX AAV25075;
AC
XX
XX 24-AUG-1999 (first entry)
DT
XX
XX TAT transduction domain peptide motif.
DE
XX
XX Anti-pathogen; fusion protein; protein transduction domain; PTD; AZT;
KM cytotoxic domain; suppressor; infection; medicament; ddi; ddC; ddT; 3TC;
KM FTC; DAPD; 1592U89; C592; acyclovir; ganciclovir; penciclovir; interferon;
KM apoptosis; virus; HIV; cytomegalovirus; CMV; herpes simplex virus; HSV-1;
KM hepatitis virus; Kaposi's sarcoma-associated herpes virus; KSHV;
KM herpes virus; yellow fever virus; flavivirus; rhinovirus; plasmoidal;
KM transduction efficiency; cytotoxin.
XX
XX
XX Unidentified.
OS
XX
XX MO9929721-A1.
PN
XX
XX 17-JUN-1999.
PD
XX
XX 10-DEC-1998; 98WO-US26358.
PF
XX
XX 20-APR-1998; 98US-0082402.
PR
XX
XX 10-DEC-1997; 97US-0069012.
PR
XX
XX (UNITV ) UNITV WASHINGTON.
PA
XX
XX Dowdy SF;
PI
XX
XX WPI; 1999-394958/33.
DR
XX

```

PT	New anti-pathogen systems, particularly for virus and plasmodium infections
XX	
PS	Claim 65; Page 34; 123pp: English.
XX	
CC	This invention describes a novel anti-pathogen system (APS) comprising a fusion protein constructed from a covalently linked protein transduction domain (PTD) and a cytotoxic domain. The APS can be used for suppressing a pathogen infection in a mammal. The method may further comprise administering a medicament e.g., AZT, ddI, ddC, dAT, 3TC, FTC, DAPD, CC 1592089, GS92, acyclovir, ganciclovir, penciclovir or an interferon. The APS can also be administered to a mammal in the presence of a pathogen to induce apoptosis in a predetermined population of cells. The products can be used for treating mammals suffering from or susceptible to a viral infection or a disease associated with a virus, e.g. HIV, cytomegalovirus (CMV), herpes simplex virus, e.g. type 1 (HSV-1) hepatitis virus, Type C (HCV). Kaposi's sarcoma-associated herpes virus (KSHV or human herpes virus 8), yellow fever virus, flavivirus or rhinovirus, or suffering from or susceptible to plasmoidal infection or a disease associated with a plasmoidal infection, e.g. P. falciparum, P. vivax, P. ovale, or P. malariae. The APS exhibits high transduction efficiency and specifically kills or injures cells infected by one or more pathogens. Formation of the cytotoxin is minimized or eliminated in uninfected cells and in infected cells that keep the pathogen inactive. The APS can be specifically tailored to kill or injure cells infected by one or more pathogen strains. This sequence represents a PAT transduction domain peptide motif which is used in the method of the invention.
XX	
SQ	Sequence    11 AA;
Query Match	84.5%; Score 49; DB 20; Length 11;
Best Local Similarity	72.7%; Pred. No. 0.12;
Matches	8; Conservative    3; Mismatches    0; Indels    0; Gaps    0;
OY	1 YGRRKRROKK 11      :::
Db	1 ygrkkrrgrrr 11
RESULT	4
AAAY05415	
ID	AAAY05415 standard; peptide; 11 AA.
XX	
AC	AAAY05415;
XX	
DJ	02-JUL-1999 (first entry)
XX	
DE	Tet peptide.
XX	
KW	BH3 domain; cell death agonist; bcl homology domain; BCL-2 family; apoptosis promoter; cancer cell; Virus infected cell; Inflammation; autoantibody producing cell; cancer; lymphoproliferative condition; arthritis; autoimmune disease; therapy.
KW	
OS	Unidentified.
XX	
PN	WO916787-A1.
PD	08-APR-1999.
XX	
XX	22-SEP-1998; 98MO-US19765.
PR	07-OCT-1997; 97US-0946039.
XX	
XX	26-SEP-1997; 97US-0060133.
PA	(UNIV ) UNIV WASHINGTON.
PI	Korsmeyer SJ.
XX	
DR	WPI; 1999-255058/21.
XX	
PT	Bcl homology domain 3 polypeptide

XX Claim 7; Page 67; 104pp; English.

PS This sequence represents a Tat peptide.

CC The invention relates to a bcl homology domain 3 (BH3 domain),

CC derived from a proapoptotic member of the BCL-2 family. The

CC BH3 polypeptide can be used in a method for promoting apoptosis in a

CC target cell, especially where the cell is a cancer cell a virus infected

CC cell or an autanctibody producing cell. The BH3 polypeptide can be used

CC in therapeutic compositions for treating disease including cancer, other

CC lymphoproliferative conditions, arthritis, inflammation, and autoimmune

CC diseases, which may result from the down regulation of cell death

CC regulation.

CC XX

SQ Sequence 11 AA;

Query Match 84.5%; Score 49; DB 20; Length 11;  
Best Local Similarity 72.7%; Pred No. 0.12;  
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 YGRRKKRROKRR 11  
|||||||:::  
Db 1 ygrkrrrrrrr 11

RESULT 5  
AAB03932 standard; peptide, 11 AA.  
ID AAB03932  
AC AAB03932;  
DT 26-FEB-2001 (first entry)  
XX  
DE TAT protein transduction domain (internalisation moiety).  
XX  
XX Modulating agent; beta-catenin; hair loss; hair growth; skin;  
KW exfoliation; Alzheimer's disease; gene transcription;  
KM cell differentiation; hearing loss; inner ear; hyperacusis; tinnitus;  
KW hair regeneration.  
XX  
OS Human immunodeficiency virus (HIV).  
XX  
PN W0200059939-A1.  
XX  
PD 12-OCT-2000.  
XX  
PF 04-APR-2000; 2000WO-US091174.  
XX  
PR 05-APR-1999; 99US-0288373.  
XX  
PA (ADHE-) ADHEREX TECHNOLOGIES INC.  
PI Blaschuk OW, Byers S, Gour BJ;  
PI WP1: 2000-679355/66.  
DR  
XX  
PT Modulating agents for inhibiting degradation of cytoplasmic  
PT beta-catenin, used for e.g. stimulating hair growth or reducing hair  
PT loss, inhibiting development of Alzheimer's disease, comprise  
PT internalization moiety and amino acid sequence  
XX  
XX  
PS Claim 4; Page 46; 49pp; English.  
XX  
CC Modulating agents for inhibiting degradation of cytoplasmic  
CC beta-catenin are described. The modulating agent comprises an  
CC internalisation moiety and one or more of an amino acid sequence  
CC SYDYS(PQ\_4)GHSH(PQ\_4)G, or a peptide analogue or peptidomimetic of  
CC the amino acid sequence. The modulating agents are useful for the  
CC manufacture of a medicament for stimulating hair growth or reducing  
CC hair loss, stimulating skin exfoliation, and inhibiting the  
CC development of Alzheimer's disease. They may also be used to  
CC increase the beta-catenin level in a cell, to stimulate activation

CC of gene transcription in a cell, and to stimulate cell  
CC differentiation. They may further be used to ameliorate hearing loss  
CC resulting from a variety of inner ear disorders, such as hyperacusis  
CC and tinnitus, through regeneration of hair cells of the inner ear.

XX  
SQ Sequence 11 AA;

Query Match 84.5%; Score 49; DB 21; Length 11;  
Best Local Similarity 72.7%; Pred. No. 0.12;  
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 YGRKKRRQKKK 11  
|||||  
Db 1 ygrkkrrqrtr 11

## RESULT 6

AAB03961  
ID AAB03961 standard; Peptide; 11 AA.

AC AAB03961;

DT 26-FEB-2001 (first entry)

DE Minimal eleven amino acids present in p18T.

XX Chimeric protein; fusion protein; FLICE like inhibitor protein;  
XX FLIP; Fas; TNF; apoptosis; caspase-8; ligand; T cell; thymocyte;  
XX tumour specific antigen; immune response; therapy; prophylaxis;  
XX diagnosis; HIV; human immunodeficiency syndrome; AIDS;  
XX acquired immune deficiency syndrome.

OS Human immunodeficiency virus.

PN W0200059935-A1.

PD 12-OCT-2000.

PF 05-APR-2000; 2000MO-US09002.

PR 05-APR-1999; 99US-0127867.

PR 06-APR-1999; 99US-0128021.

PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

PA (PAYA/) PAYA C.

PI (ALGE/) ALGECIRAS-SCHMINICH A.

PI Paya C, Algeciras-schminich A;

DR WPI: 2000-664988/54.

PT Fusion polypeptide useful for inhibiting ligand-induced apoptosis.

PT comprises portion of anti-apoptotic polypeptide linked to a transport

PT group

PS Disclosure: Page 30; 89pp; English.

XX A chimeric group or fusion peptide which comprises a portion of an  
XX anti-apoptotic polypeptide which inhibits apoptosis of lymphocytes in  
XX combination with a transport group is described. The transport group  
XX is capable of transporting the chimeric group or fusion peptide  
XX across the cell membrane. The anti-apoptotic polypeptide is FLICE-like  
XX inhibitor protein (FLIP) which inhibits Fas and TNF mediated apoptosis  
XX by inhibiting binding of Caspase-8 to the Fas receptor complex, thus  
XX shutting off the downstream Fas signalling pathway. The chimeric group  
XX and fusion peptide are useful for inhibiting ligand-induced apoptosis  
XX by bringing them into contact with T cells. The chimeric group is  
XX useful for expanding T cells in vitro e.g. T cells specific for  
XX particular antigens such as tumour-specific antigen, for enhancing  
XX immune response and to inhibit the apoptosis of chronically activated  
XX T cells e.g. activated CD4+ T cells in HIV infected patients. The  
XX chimeric group is also useful for therapeutic, prophylactic or

CC diagnosis of intracellular delivery of small molecules and  
CC macromolecules such as anti-apoptotic polypeptides and nucleic  
CC acids encoding such polypeptides. Two primers (AA454297, AA454298) were  
CC used to amplify the FLIP cDNA for subcloning into the XhoI-NcoI  
CC site of the pHA-TAT vector which contains the N-terminal protein  
CC transduction domain from the human immunodeficiency virus tat  
CC protein. Tat is a preferred transport moiety.

XX  
SQ Sequence 11 AA;

Query Match 84.5%; Score 49; DB 21; Length 11;  
Best Local Similarity 72.7%; Pred. No. 0.12;  
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 YGRKKRRQKKK 11  
|||||  
Db 1 ygrkkrrqrtr 11

## RESULT 7

AAB27088  
ID AAB27088 standard; Peptide; 11 AA.

AC AAB27088;

DT 15-FEB-2001 (first entry)

DE Beta-catenin derived internalisation moiety SEQ ID NO: 75.

XX Beta-catenin; cadherin-mediated intercellular adhesion;  
XX cell differentiation; modulating agent; hair loss; skin exfoliation;  
XX internalisation moiety; flanking sequence; transcription; hearing loss.

OS Human immunodeficiency virus.

PN W0200053632-A1.

PD 14-SEP-2000.

PF 07-MAR-2000; 2000MO-CA00222.

PR 09-MAR-1999; 99US-0265107.

PA (UYMC-) UNIV MCGILL.

PA Blaschuk OM, Gour BJ;

DR WPI: 2000-594308/56.

PT Stimulating beta-catenin mediated gene expression, cellular  
PT differentiation and hair growth, involves contacting cells with  
PT modulating agent capable of inhibiting interaction between alpha and  
PT beta catenin

PS Disclosure: Page 19; 77pp; English.

XX The present invention is concerned with methods of modulating the amount  
XX of free beta-catenin in the cell, and methods of stimulating the  
XX expression of genes involved in cellular differentiation, the  
XX transcription of which is under the control of beta-catenin. The  
XX peptides given in AAB27053-B27088, AAB27284-B27300 and AAB27330-B27351  
XX can be used as modulating agents which interrupt the interaction between  
XX alpha and beta catenin, causing increased levels of the latter and  
XX stimulating the activation of beta-catenin mediated transcription. This  
XX can be used to stimulate cell differentiation, which can then be used to  
XX promote hair growth and skin exfoliation. This latter is particularly  
XX useful in the improvement of photodamaged skin and to minimise wrinkles.  
XX The modulating peptide can also be used to reduce hearing loss resulting  
XX from inner ear disorders such as hyperacusis and tinnitus.

XX  
SQ Sequence 11 AA;



```

OY      1 YGRRKRRQKK 11
DB      1 YGRRKRRQRR 11

RESULT 10
AAB09907
ID      AAB09907 standard; peptide; 11 AA.
XX
AC      AAB09907;
XX
DT      06-NOV-2000 (first entry)
XX
DE      HIV tat protein transduction domain.
XX
XX      Targeting signal; MUC-1; immunosuppression; autoimmune disorder;
XX      immune disorder; inflammatory disorder.
XX
OS      Human immunodeficiency virus.
XX
PN      WO200034468-A2.
XX
PD      15-JUN-2000.
XX
PF      09-DEC-1999; 99WO-US29016.
XX
PR      11-DEC-1998; 98US-0111973.
XX
PA      (BIOM-) BIOMIRA INC.
XX
PI      Agrawal B, Longenecker BM;
XX
DR      WPI; 2000-423418/36.
XX
PT      Use of agent capable of intracellularly inhibiting mucin MUC-1 for
PT      inducing T-cell-based immunosuppression and for treating autoimmune
PT      disorders, transplant rejection and inflammatory disorders -
XX
PS      Disclosure: Page 15; 51pp; English.
XX
CC      The present sequence is the tat protein transduction domain from HIV,
CC      which can be used as a targeting signal. It can be used to internalise
CC      sequences, such as MUC-1 antagonists, within the cell. MUC-1 is a
CC      immunosuppressor, and antagonists act to reduce overactive immune
CC      responses. Thus, the peptide can be used to treat inflammatory disorders
CC      such as rheumatoid arthritis, psoriasis, allergic contact dermatitis and
CC      ankylosing spondylitis, autoimmune disorders including myasthenia gravis,
CC      systemic lupus erythematosus, polyarteritis nodosa, Goodpastures
CC      syndrome, Isopathic thrombocytopenic purpura, autoimmune haemolytic
CC      anaemia, Graves' disease, rheumatic fever, pernicious anaemia,
CC      insulin-resistant diabetes mellitus, bullous pemphigoid, pemphigus
CC      vulgaris, viral myocarditis, autoimmune thyroiditis, male infertility,
CC      sarcoidosis, allergic encephalomyelitis, multiple sclerosis, Sjogrens
CC      disease, Reiter's disease, Cellac disease, sympathetic ophthalmia and
CC      primary biliary cirrhosis, immune disorders, graft versus host disease
CC      and transplant rejection.
XX
SQ      Sequence 11 AA;

Query Match      84.5%; Score 49; DB 21; Length 11;
Best Local Similarity 72.7%; Pred. No. 0.12;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

AC      AAY93542;
XX
DT      25-SEP-2000 (first entry)
XX
DE      Amino acid sequence of a synthetic protein transduction domain.
XX
XX      Protein transduction system; protein transduction domain;
XX      cytotoxic domain; pathogen infection; retroviral infection;
XX      plasmidial infection; cancer; prostate cancer.
XX
OS      Synthetic.
XX
PN      WO200034308-A2.
XX
PD      15-JUN-2000.
XX
PF      10-DEC-1999; 99WO-US29289.
XX
PR      10-DEC-1998; 98US-0111701.
XX
PA      (UNIV ) UNIV WASHINGTON.
XX
PI      Dowdy SF;
XX
DR      WPI; 2000-431269/37.
XX
PT      Protein transduction system for treating cancer and pathogenic
PT      infections has a fusion protein comprising a protein transduction
PT      domain covalently linked to a cytotoxic domain -
XX
PS      Example 10; Page 71; 127pp; English.
XX
CC      AAY93542-51 represent synthetic protein transduction domains, which
CC      are used in the protein transduction system of the invention. The
CC      specification describes a protein transduction system, linked protein
CC      a fusion protein. This fusion protein has a covalently linked protein
CC      transduction domain and cytotoxic domain. The system is useful for
CC      treating pathogen infection in mammals. Infections such as those
CC      caused by CMV, HSV-1, HCV, KSHV, yellow fever virus, flavivirus or
CC      rhinovirus, retroviral infections such as HIV-1, HIV-2, HTLV-3 and/or
CC      LAV, plasmidial infections associated with P.faciaparum, P.vivax,
CC      P.ovale, P.malariae. It is also useful for treating cancer, especially
CC      prostate cancer.
XX
SQ      Sequence 11 AA;

Query Match      84.5%; Score 49; DB 21; Length 11;
Best Local Similarity 72.7%; Pred. No. 0.12;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

RESULT 12
AAY71015
ID      AAY71015 standard; peptide; 11 AA.
XX
AC      AAY71015;
XX
DT      29-AUG-2000 (first entry)
XX
DE      Human immunodeficiency virus TAR peptide.
XX
XX      Phospholamban; PLB; sarcoplasmic reticulum Ca 2+ ATPase; SERCA2a; HIV;
XX      TAR protein; penetratin; transport peptide; cardiomyocyte inhibitor;
XX      cargo peptide; contractilin; cardiac contractility; cardiant; treatment;
XX      cardiac disease; heart failure; myocardial dysfunction.
XX
OS      Human immunodeficiency virus.
XX

```



PN	WO260025804-A2.
XX	
PD	11-MAY-2000.
XX	
PF	02-NOV-1999; 99WO-US25692.
XX	
PR	02-NOV-1998; 98US-0106718.
PR	27-JUL-1999; 99US-0145883.
XX	
PA	(REGC ) UNIV CALIFORNIA.
XX	
PI	Chien K, Dillman W, Minamisawa S, He H, Hoshijima M, Meyer M;
PI	Scott C, Wang Y, Silverman GJ;
XX	
XR	WPI; 2000-365393/31.

PT Treating cardiac diseases, e.g. heart failure or myocardial dysfunction  
PT comprises enhancing cardiac contractility by inhibiting interaction  
PT between phospholamban and sarcoplasmic reticulum calcium 2+ adenosine  
PT triphosphatase -  
XX  
PS Example 5; Page 54; 56pp; English.

The patent discloses a method for the treatment of heart failure, using small peptide complexes and recombinant proteins, that induces phospholamban (PLB) deficiency and inhibits the interaction between PLB and sarcoplasmic reticulum Ca<sup>2+</sup> ATPase (SERCA2a) within cardiomyocytes. The peptide complex comprises of transport peptide like penetratin and cargo peptide selected from mutant PLB, native PLB or antibody against PLB protein (contractilin). Penetratin-PLB peptide functions as a dominant negative inhibitor of PLB-SERCA2a interaction, enhances cardiactility and reduces blood pressure. This method is useful for the treatment of cardiac disease e.g., heart failure and myocardial dysfunction. The present sequence is the amino terminal, 11-amino acid protein transduction domain of the denuated human immunodeficiency virus (HIV) Tat protein. It belongs to the penetratin class of peptides with translocating properties having the ability to carry hydrophilic compounds across the plasma membrane. It is used in the construction of recombinant PLB peptides (e.g., TAT-PLB).

50 Sequence 11 AA;

Query Match	84.5%;	Score 49;	DB 21;	Length 11;
Best Local Similarity	72.7%;	Pred. No. 0.12;		
Matches	8;	Conservative	3;	Mismatches 0;
			Indels	0;
			Gaps	0

```
OY 1 YGRKKRRQKK 11
      |||||:::
Db 1 YGRKKrrrr 11
```

RESULT 13  
AAM09937

ID AU09932 standard; peptide; 11 AA.

AC AAU09932;

DT 15-JAN-2002 (first entry)

# DE Human immunodeficiency virus (HIV) tat partial peptide sequence

KW Human; fibroblast growth factor-like; EGF-L; HIV tat protein;

KW protein transduction domain; PDT; human immunodeficiency virus.

OS Human immunodeficiency virus type 1.

PN W0200168854-A2.

PD 20-SEP-2001

PF 13-MAR-2001; 2001WO-US08013.

PR 13-MAR-2000; 2000US-188786P.

PA (AMGE-) AMGEN INC

PI Jing S, Bass MB,

DR WPI; 2001-596910/67

PT New fibroblast growth factor-like polypeptide and polynucleotide for

PT conditions involving the central nervous system, teeth, heart, liver or

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841

The present invention relates to new isolated fibroblast growth factor-like (FGF-L) nucleic acid molecules and polypeptides. The FGF-L polypeptide is useful for determining whether a compound inhibits FGF-L polypeptide, and also for identifying a compound that binds to the polypeptide. The FGF-L polypeptides of the invention are useful for treating, preventing or ameliorating a medical condition or an FGF-L polypeptide-related disease, condition or disorder such as wound healing disorders, ulcers, gut disorders, lung disorders, liver disorders such as hepatitis and diabetes. The invention is also useful for diagnosing a pathological condition or susceptibility to a pathological condition in a subject and is useful for modulating levels of FGF-L in an animal. Other uses are detecting or quantifying the amount of FGF-L polypeptide and for identifying or developing novel agonists and antagonists of the FGF-L polypeptide signaling pathway which are useful for treating one or more diseases or disorders, and also as an immunogen for producing antibodies for *in vivo* imaging. The present sequence represents an 11 amino acid sequence from the human immunodeficiency virus (HIV) tat protein that is termed the protein transduction domain or TAT PRD.

Sequence 11 AA;

Query Match	84.58; Score 49; DB 22; Length 11;
Query Match	84.58; Score 49; DB 22; Length 11;

Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 YGRRKRQKK 11

D5 1 ygrkkrqr 11

## RESULT 14

ID AAU09812

AC AAU09812;

DT 27-FEB-2002 (first entry)

DE HIV-1 tat protein transduction domain (TAT PDT)

KM fibroblast growth factor receptor-like protein; FGFR-L; anorectic;

KW Immunomodulator; antinflammatory; haematopoietic disorder; osteoporosis;

KW hypercalcaemia; acute glomerulonephritis; chronic glomerulonephritis

KW human immunodeficiency virus; HIV-1; tat protein

OS Human immunodeficiency virus type 1.

PN WO200170977-A2.

PD 27-SEP-2001

PF 22-MAR-2001; 2001WO-US09073

PR 22-MAR-2000; 2000US-191379P

XX (AMGE-) AMGEN INC.  
PA (SARI/) SARIS C M.  
PA (MUSX/) MU S X.  
PA (XIAM/) XIA M.  
PA (BOON/) BOONE T C.  
PA (COVE/) COVER T.  
PI Saris CM, Mu SX, Xia M, Boone TC, Covey T;  
XX WPI: 2001-626128/72.  
XX Novel nucleic acid encoding fibroblast growth factor receptor-like  
PT polypeptides, useful for treating hematopoietic disorder, osteoporosis,  
PT Paget's disease, glomerulonephritis, cancer, diabetes, obesity and  
PT cachexia -  
XX  
XX  
PS Disclosure; Page 59; 163pp; English.  
XX  
XX The invention relates to a novel isolated fibroblast growth factor  
CC receptor-like (FGFR-L) polypeptide (I). (I) and the nucleic acid (II)  
CC encoding (I) are useful for treating, preventing or ameliorating  
CC a medical condition including hematopoietic disorder, osteoporosis,  
CC osteogenesis imperfecta, Paget's disease, periodontal disease,  
CC hypercalcaemia, acute glomerulonephritis, chronic glomerulonephritis,  
CC cancer, diabetes, obesity and cachexia. (I) is also useful for  
CC identifying a compound which binds to FGFR-L polypeptide, by contacting  
CC (I) with a compound, determining the extent of binding of the FGFR-L  
CC polypeptide to the compound, and determining the activity of the  
CC polypeptide when bound to the compound. (II) is useful for modulating  
CC levels of a polypeptide in an animal. A transgenic animal comprising (I)  
CC is useful for determining whether a compound inhibits FGFR-L polypeptide  
CC activity or FGFR-L polypeptide production, by exposing the transgenic  
CC animal to the compound and measuring FGFR-L polypeptide or production in  
CC the animal. (II) is useful for mapping the locations of FGFR-L gene and  
CC related genes on chromosomes, as hybridisation probes in diagnostic  
CC assays to test for the presence of an FGFR-L nucleic acid molecule in  
CC mammalian tissue or bodily fluid samples, in gene therapy, and as tools  
CC for isolating corresponding FGFR-L polypeptide genes. (I) is useful as  
CC immunogen, and for cloning FGFR-L polypeptide ligands using an  
CC expression cloning strategy. The present sequence represents the  
CC amino acid sequence of human immunodeficiency virus (HIV-1) tat protein  
CC transduction domain (TAT PDT) as described in the invention.  
XX  
XX  
SQ Sequence 11 AA;  
Query Match 84.5%; Score 49; DB 22; Length 11;  
Best Local Similarity 72.7%; Pred. No. 0.12;  
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
OY 1 YGRRKRQKK 11  
Db 1 YGRKKRQRR 11  
RESULT 15  
AAE13064  
ID AAE13064 standard; peptide; 11 AA.  
XX  
XX AAE13064:  
AC  
XX  
XX 28-JAN-2002 (first entry)  
DE  
XX  
XX Protein transduction domain of HIV tat protein.  
KW Human: CD20/19F-receptor like protein; immunoglobulin E: agp-96614-a1;  
KW agp-69406-a1; cancer; abnormal cell proliferation; autoimmune disease;  
KW ovarian cancer; brain cancer; arteriosclerosis; vascular restenosis;  
KW rheumatoid arthritis; multiple sclerosis; allergy; dermatitis; asthma;  
KW reproductive disease; diabetes; transplant rejection; endometriosis;  
KW infertility; gene therapy; protein transduction domain; HIV;  
KW human immunodeficiency virus; TAT PDT.

XX  
OS Human immunodeficiency virus.  
XX  
XX WO200174903-A2.  
PN  
XX  
XX 11-OCV-2001.  
PD  
XX  
XX 29-MAR-2001; 2001WO-US10048.  
PE  
XX  
XX 30-MAR-2000; 2000US-193728P.  
PR 27-NOV-2000; 2000US-0723258.  
XX  
XX (AMGE-) AMGEN INC.  
PA  
XX  
XX Welcher AA, Calzone FJ;  
PI  
XX  
XX WPI: 2001-662968/76.  
DR  
XX  
XX Novel CD20/19F-receptor like polypeptides and polynucleotides,  
PT antagonists and antibodies of the polypeptide useful for treating  
PT ameliorating or preventing diseases associated with the polypeptide  
PT e.g. cancer, asthma -  
XX  
XX  
PS Disclosure; Page 82; 145pp; English.  
XX  
XX The invention relates to human CD20/immunoglobulin E (19F)-receptor  
CC like polypeptides designated as agp-96614-a1 and agp-69406-a1 and  
CC nucleic acid molecules encoding such polypeptides. Polypeptides of  
CC the invention are useful for treating, preventing or ameliorating  
CC a disease, condition, or disorder which includes cancer such as  
CC brain cancer, ovarian cancer; abnormal cell proliferation such as  
CC arteriosclerosis, vascular restenosis; pathology from allergens  
CC such as allergies, asthma, dermatitis; dysfunction of immune system  
CC such as rheumatoid arthritis, autoimmune disease, multiple sclerosis,  
CC diabetes; transplant rejection and reproductive diseases such as  
CC infertility, preterm labour and delivery, endometriosis etc. They  
CC are also useful for identifying antagonists and as immunogens, for  
CC raising antibodies which may also be used to prevent, treat or  
CC diagnose a number of diseases and disorders. Polynucleotides of the  
CC invention are used to map the location of CD20/19F-receptor like  
CC gene and related genes on chromosomes and as hybridisation probes.  
CC They are also useful in gene therapy. The present sequence is protein  
CC transduction domain (TAT PDT) of human immunodeficiency virus (HIV)  
CC tat protein. This sequence is used to internalise protein of the  
CC invention into a cell.  
XX  
XX  
SQ Sequence 11 AA;  
Query Match 84.5%; Score 49; DB 22; Length 11;  
Best Local Similarity 72.7%; Pred. No. 0.12;  
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
OY 1 YGRRKRQKK 11  
Db 1 YGRKKRQRR 11  
Search completed: July 15, 2002, 13:47:08  
Job time: 207 sec



```
RESULT 2
US-08-924-695A-54
; Sequence 54, Application US/08924695A
; Patent No. 5998983
; GENERAL INFORMATION:
; APPLICANT: KORSMEYER, STANLEY J.
; TITLE OF INVENTION: BH3 INTERACTING DOMAIN DEATH AGONIST
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/924,695A
; FILING DATE: 09-SEP-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 971798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: Peptide
; US-08-924-695A-54

Query Match      84.5%; Score 49; DB 2; Length 11;
Best Local Similarity 72.7%; Pred. No. 0.067;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 YGKRRRQKKK 11
Db      1 YGKRRRQRRR 11

RESULT 3
US-09-208-966-2
; Sequence 2, Application US/09208966
; Patent No. 6221355
; GENERAL INFORMATION:
; APPLICANT: Dowdy, Steven F.
; TITLE OF INVENTION: ANTI-PATHOGEN SYSTEM AND METHODS OF USE THEREOF
; FILE REFERENCE: 48881/1742
; CURRENT APPLICATION NUMBER: US/09/208,966
; CURRENT FILING DATE: 1998-12-10
; EARLIER APPLICATION NUMBER: 60/082,402
; EARLIER FILING DATE: 1998-04-20
; EARLIER APPLICATION NUMBER: 60/069,012
; EARLIER FILING DATE: 1997-12-10
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 2
; LENGTH: 11
; TYPE: PRT
; ORGANISM: human
US-09-208-966-2
```

```
Query Match      84.5%; Score 49; DB 4; Length 11;
Best Local Similarity 72.7%; Pred. No. 0.067;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 YGKRRRQKKK 11
Db      1 YGKRRRQRRR 11

RESULT 4
US-09-296-089-37
; Sequence 37, Application US/09296089
; Patent No. 6303576
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Byers, Stephen
; APPLICANT: Cour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING
; FILE REFERENCE: 100086,411
; CURRENT APPLICATION NUMBER: US/09/296,089
; CURRENT FILING DATE: 1999-04-21
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 37
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Human Immunodeficiency virus
US-09-296-089-37
```

```
Query Match      84.5%; Score 49; DB 4; Length 11;
Best Local Similarity 72.7%; Pred. No. 0.067;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 YGKRRRQKKK 11
Db      1 YGKRRRQRRR 11

RESULT 5
US-09-336-093-7
; Sequence 7, Application US/09336093A
; Patent No. 6348185
; GENERAL INFORMATION:
; APPLICANT: Washington University School of Medicine
; TITLE OF INVENTION: MEMBRANE-PERMEANT PEPTIDE COMPLEXES FOR MEDICAL
; FILE REFERENCE: WSHU 2001
; CURRENT APPLICATION NUMBER: US/09/336,093A
; CURRENT FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 7
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: TAT protein
US-09-336-093-7
```

```
Query Match      72.4%; Score 42; DB 4; Length 10;
Best Local Similarity 70.0%; Pred. No. 0.72;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      2 GKRRRQKKK 11
Db      1 GKRRRQRRR 10
```

RESULT 6

```

US-09-208-966-16
; Sequence 16, Application US/09208966
; Patent No. 6221355
; GENERAL INFORMATION:
; APPLICANT: Dowdy, Steven F.
; TITLE OF INVENTION: ANTI-PATHOGEN SYSTEM AND METHODS OF USE THEREOF
; FILE REFERENCE: 48881/1742
; CURRENT APPLICATION NUMBER: US/09/208,966
; CURRENT FILING DATE: 1998-12-10
; EARLIER APPLICATION NUMBER: 60/082,402
; EARLIER FILING DATE: 1998-04-20
; EARLIER APPLICATION NUMBER: 60/069,012
; EARLIER FILING DATE: 1997-12-10
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 11
; TYPE: PRT
; ORGANISM: human
US-09-208-966-16

```

```

Query Match          72.4%; Score 42; DB 4; Length 11;
Best Local Similarity 70.0%; Pred. No. 0.78;
Matches 7; Conservative 3; Mismatches 0; Gaps 0;

```

```

QY      2 GRRKKRQKKK 11
        |||||
Db      1 GRRKKRQRRR 10

```

```

RESULT 7
US-09-208-966-52
; Sequence 52, Application US/09208966
; Patent No. 6221355
; GENERAL INFORMATION:
; APPLICANT: Dowdy, Steven F.
; TITLE OF INVENTION: ANTI-PATHOGEN SYSTEM AND METHODS OF USE THEREOF
; FILE REFERENCE: 48881/1742
; CURRENT APPLICATION NUMBER: US/09/208,966
; CURRENT FILING DATE: 1998-12-10
; EARLIER APPLICATION NUMBER: 60/082,402
; EARLIER FILING DATE: 1998-04-20
; EARLIER APPLICATION NUMBER: 60/069,012
; EARLIER FILING DATE: 1997-12-10
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 52
; LENGTH: 11
; TYPE: PRT
; ORGANISM: human
US-09-208-966-52

```

```

Query Match          72.4%; Score 42; DB 4; Length 11;
Best Local Similarity 70.0%; Pred. No. 0.78;
Matches 7; Conservative 3; Mismatches 0; Gaps 0;

```

```

QY      2 GRRKKRQKKK 11
        |||||
Db      2 GRRKKRQRRR 11

```

```

RESULT 8
US-07-694-983-15
; Sequence 15, Application US/07694983
; Patent No. 5432260
; GENERAL INFORMATION:
; APPLICANT: Stahl, Philip D.
; TITLE OF INVENTION: HIGH AFFINITY MANNOS RECEPTOR
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Irell & Manella
; STREET: 545 Middlefield Road, Suite 200
; CITY: Menlo Park
; STATE: California
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/694,983
; FILING DATE: 19910503
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 9500-0039.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-327-7250
; TELEFAX: 415-327-2951
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1
; OTHER INFORMATION: /label= Ac-
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 11
; OTHER INFORMATION: /label= -NH2
US-07-694-983-15

```

```

Query Match          63.8%; Score 37; DB 1; Length 11;
Best Local Similarity 54.5%; Pred. No. 4.5;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 YGKKRRQKKK 11
        | :|||:||||
Db      1 YKKKKKKKKK 11

```

```

RESULT 9
PCT-US95-06077-17
; Sequence 17, Application PC/TUS9506077
; GENERAL INFORMATION:
; APPLICANT: Immunobiology Research, Institute Inc.
; TITLE OF INVENTION: Vaccine Interdiction of Extracellular
; TITLE OF INVENTION: Transactivating Proteins of Human Immunodeficiency Virus
; TITLE OF INVENTION: and Other Chronically Infecting Viruses Employing Similar
; TITLE OF INVENTION: InterCellular Transactivating Strategies
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: PCT/US95/06077  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/247,991  
FILING DATE: 23-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: IR144PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
PCT-US95-06077-17

Query Match 62.1%; Score 36; DB 5; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.7e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 GRRRRQ 8  
DB 1 GRRRRQ 7

RESULT 10  
US-08-218-608-3  
Sequence 3, Application US/08218608  
Patent No. 5607859  
GENERAL INFORMATION:  
APPLICANT: BIEMANN, KLAUS  
TITLE OF INVENTION: METHODS AND PRODUCTS FOR MASS  
TITLE OF INVENTION: SPECTROMETRIC MOLECULAR WEIGHT DETERMINATION OF POLYIONIC  
TITLE OF INVENTION: ANALYTES EMPLOYING POLYIONIC REAGENTS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
STREET: 600 ATLANTIC AVENUE  
CITY: BOSTON  
STATE: MA  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/218,608  
FILING DATE: 28-MAR-1994  
CLASSIFICATION: 436  
ATTORNEY/AGENT INFORMATION:  
NAME: GATES, EDWARD R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: M0656/7013  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO

ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: NONE (SYNTHETIC PEPTIDE)  
US-08-218-608-3

Query Match 62.1%; Score 36; DB 1; Length 9;  
Best Local Similarity 66.7%; Pred. No. 1.7e+05;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 3 RRRRRQ 11  
DB 1 RRRRRQ 9

RESULT 11  
US-08-357-056-1  
Sequence 1, Application US/08357056  
Patent No. 5646120  
GENERAL INFORMATION:  
APPLICANT: SUMNER-SMITH, Martin  
APPLICANT: BARNETT, Richard W.  
APPLICANT: REID, Lorne S.  
APPLICANT: SONENBERG, Nahum  
TITLE OF INVENTION: PEPTIDE-BASED INHIBITORS OF HIV  
TITLE OF INVENTION: REPLICATION  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/357,056  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/779,735  
FILING DATE: 23-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16777/140 ALLE  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-357-056-1

Query Match 62.1%; Score 36; DB 1; Length 9;  
Best Local Similarity 66.7%; Pred. No. 1.7e+05;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 3 RRRRRQ 11  
DB 1 RRRRRQ 9

RESULT 12

```
US-08-475-583-1
; Sequence 1, Application US/08475583
; Patent No. 5786531
; GENERAL INFORMATION:
; APPLICANT: SOMNER-SMITH, Martin
; APPLICANT: BARNETT, Richard W.
; APPLICANT: REID, Lorne S.
; APPLICANT: SONENBERG, Nahum
; TITLE OF INVENTION: PEPTIDE-BASED INHIBITORS OF HIV
; TITLE OF INVENTION: REPLICATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lerner
; STREET: 3000 K Street, Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,583
; FILING DATE: 07-JUN-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/779,735
; FILING DATE: 23-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16777/277/ALLI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-475-583-1

Query Match          62.1%; Score 36; DB 1; Length 9;
Best Local Similarity 66.7%; Pred. No. 1.7e+05;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      3 RKKRRQKK 11
Db      1 RKKRRQRR 9

RESULT 13
US-08-442-461D-19
; Sequence 19, Application US/08442461D
; Patent No. 5834184
; GENERAL INFORMATION:
; APPLICANT: Harada, Kazuo
; APPLICANT: Martin, Shelley S.
; APPLICANT: Frankel, Alan
; TITLE OF INVENTION: In Vivo Selection of RNA-Binding
; Peptides
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
```

```
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/442,461D
; FILING DATE: 17-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebeschuetz, Joe
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 02307U-060500US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-442-461D-19

Query Match          62.1%; Score 36; DB 2; Length 9;
Best Local Similarity 66.7%; Pred. No. 1.7e+05;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      3 RKKRRQKK 11
Db      1 RKKRRQRR 9

RESULT 14
US-08-893-853-4
; Sequence 4, Application US/08893853
; Patent No. 5891094
; GENERAL INFORMATION:
; APPLICANT: Goldstein, Gideon
; TITLE OF INVENTION: Methods and Compositions for Impairing
; Multiplication of HIV-1
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr., P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,853
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GGP2USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
```

MOLECULE TYPE: peptide  
US-08-893-853-4

Job time: 118 sec

Query Match 62.1%; Score 36; DB 2; Length 9;  
Best Local Similarity 66.7%; Pred. No. 1.7e+05;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 3 RKKRRKKK 11  
|||||:  
DB 1 RKKRRQRRR 9

RESULT 15  
US-08-564-164A-10  
; Sequence 10, Application US/08564164A  
; Patent No. 6153947  
; GENERAL INFORMATION:  
; APPLICANT: Schweighoffer, Fabien  
; APPLICANT: Tocque, Bruno  
; TITLE OF INVENTION: Intracellular Binding Proteins and Use  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Arcola Road, 3C43  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426-0107  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/564,164A  
; FILING DATE: 28-DEC-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/FR94/00714  
; FILING DATE: 15-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 93/07241  
; FILING DATE: 16-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sawitzky, Martin F.  
; REGISTRATION NUMBER: 29,699  
; REFERENCE/DOCKET NUMBER: ST93030-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (610)454-3816  
; TELEFAX: (610)454-3808  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-564-164A-10

Query Match 62.1%; Score 36; DB 4; Length 9;  
Best Local Similarity 66.7%; Pred. No. 1.7e+05;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 3 RKKRRKKK 11  
|||||:  
DB 1 RKKRRQRRR 9

Search completed: July 15, 2002, 13:47:49



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model.

Run on: July 15, 2002, 13:50:01 ; Search time 25.22 Seconds

(without alignments)  
110,491 Million cell updates/sec

Title: US-09-712-819a-9

Perfect score: 144

Sequence: 1 YEKRRRRRRSGSGTDFLTITSSIQAD 29

Scoring table: BLOSUM62

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 5929

Minimum DB seq length: 0

Maximum DB seq length: 29

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR\_71:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	29.2	23	2	SA3765 ribosomal protein
2	39	27.1	29	2	S08555 ribosomal protein
3	38	26.4	27	1	SRAPAS protamine A - steel
4	35	24.3	21	2	PR0082 sperm chromatin pr
5	35	24.3	25	1	JC4278 ribosomal protein
6	35	24.3	25	1	JQ1617 ribosomal protein
7	35	24.3	25	2	JC4685 ribosomal protein
8	35	24.3	25	2	S38425 ribosomal protein
9	35	24.3	25	2	T06233 ribosomal protein
10	35	24.3	25	2	T49214 ribosomal protein
11	34	23.6	27	1	SRAPC protamine B - Russ
12	33	22.9	15	2	A58304 heterogeneus ribo
13	33	22.9	25	2	T87719 ribosomal protein
14	32	22.2	29	2	S35924 T-cell receptor ga
15	31	21.5	21	2	C49042 Ig heavy chain V r
16	31	21.5	22	2	B49042 Ig heavy chain V r
17	31	21.5	22	2	I41299 T-cell receptor be
18	31	21.5	25	1	R6947B ribosomal protein
19	31	21.5	28	2	I59477 antigen, T-cell re
20	30	20.6	17	2	S26744 Ig heavy chain J r
21	29	20.1	12	2	S28544 T-cell receptor b
22	29	20.1	20	2	S09720 2S albumin small c
23	29	20.1	22	2	C64330 C64330
24	29	20.1	28	2	I46921 gene Bc1a protein
25	29	20.1	28	2	S26254 rel protein - chic
26	28	19.4	10	2	PH0946 T-cell receptor be
27	28	19.4	16	2	A35552 caldesmon - turkey
28	28	19.4	19	2	A33361 CAMP-regulated pho
29	28	19.4	27	2	S09304 homeotic protein c

30	28	19.4	28	2	I32529 Ig lambda chain V
31	27.5	19.1	23	2	S47192 T-cell receptor J-
32	27	18.8	12	2	S26546 T-cell receptor be
33	27	18.8	13	2	A53608 neurotensin - guin
34	27	18.8	14	2	PH0915 T-cell receptor be
35	27	18.8	15	2	A49480 major immunophilin
36	27	18.8	17	2	T24687 T-cell receptor be
37	27	18.8	18	2	PS0387 platelet-derived g
38	27	18.8	26	2	T14041 MADH dehydrogenase
39	26	18.1	10	2	E60787 sperm-activating p
40	26	18.1	13	1	UNBO neurotensin - bovi
41	26	18.1	15	2	I53284 T-cell receptor al
42	26	18.1	20	2	S06466 T-cell receptor al
43	26	18.1	21	2	PS0146 histone H1 - sea u
44	26	18.1	22	2	A28524 diamniopionate
45	26	18.1	23	2	I39681 exeg protein - Aer

## ALIGNMENTS

## RESULT 1

S43765 ribosomal protein S19 - evening primrose mitochondrion (fragment)

C:Species: mitochondrion Oenothera villaricae (evening primrose)

C>Date: 10-Dec-1994 #sequence, revision 12-May-1995 #text\_change 13-Aug-1999

C:Accession: S43765; S30542

R:Bock, H.; Brennicke, A.; Schuster, W.

Plant Mol. Biol. 24, 811-818, 1994

A:Title: Rps3 and rpl16 genes do not overlap in Oenothera mitochondria: GTC as a pote

A:Reference number: S43765; MUID:94250844

A:Accession: S43765

A:Molecule type: DNA

A:Residues: 1-23 <BOC>

A:Cross-references: EMBL:X69140; NID:g13188; PIDN:CAA48893.1; PID:g13189

A:Note: the source is designated as Oenothera berteriana

C:Genetics:

A:Gene: RPS19

A:Genome: mitochondrion

C:Superfamily: Escherichia coli ribosomal protein S19

C:Keywords: mitochondrion; protein biosynthesis; ribosome

Query Match

Best Local Similarity 29.2%, Score 42; DB 2; Length 23;

Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY: 1 YEKRRRRRRSGSG 15

DB: 3 FTKKRRRRSRTNIGSG 17

## RESULT 2

S08555 ribosomal protein L15 [validated] - Halobacterium salinarum (fragment)

N:Alternate names: ribosomal protein HL16

C:Species: Halobacterium salinarum

C>Date: 31-Dec-1990 #sequence, revision 31-Dec-1990 #text\_change 21-Jul-2000

C:Accession: S08555

R:Matheson, A.T.; Yaguchi, M.; Christensen, P.; Rollin, C.F.; Hasnain, S.

Can. J. Biochem. Cell Biol. 62, 426-433, 1984

A:Title: Purification, properties, and N-terminal amino acid sequence of certain 50S

A:Reference number: S07437; MUID:84282108

A:Accession: S08555

A:Molecule type: protein

A:Residues: 1-29 <MAV>

A:Note: the source is designated as Halobacterium cutirubrum

C:Superfamily: rat ribosomal protein L27a

C:Keywords: protein biosynthesis; ribosome

Query Match

27.1%, Score 39; DB 2; Length 29;

Best Local Similarity 66.7%; Pred. No. 93;  
Matches 10; Conservative 0; Mismatches 3; Indels 2; Gaps 1;  
QY 4 KKRNR--RNSGCT 16  
| | | | |  
| | | | |  
Db 3 KKRNRGSRTHGCT 17

RESULT 3  
SRAPAS  
protamine A - stellate sturgeon  
N:Alternate names: stellin  
C:Species: Acipenser stellatus (stellate sturgeon)  
C>Date: 28-Feb-1980 #sequence\_revision 12-Aug-1981 #text\_change 16-Feb-1997  
C:Accession: A02666  
R:Yulikova, E.P.; Rydin, V.K.; Sillaev, A.B.  
Bioorg. Khim. 5, 5-10, 1979  
A>Title: The primary structure of stellin A.  
A:Reference number: A02666  
A:Accession: A02666  
A:Molecule type: protein  
A:Residues: 1-27 <YUL>  
A:Note: article in Russian with English abstract  
C:Superfamily: protamine Y2  
C:Keywords: chromosomal protein; DNA binding; spermatogenesis

Query Match 26.4%; Score 38; DB 1; Length 27;  
Best Local Similarity 70.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 KKRNR--RNSGCT 13  
| | | | |  
| | | | |  
Db 13 KKRNR--RNSGCT 22

RESULT 4  
PN0082  
sperm chromatin protein 12-2 - Argentinian shortfin squid (fragment)  
C:Species: Illex argentinus (Argentinian shortfin squid)  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
R:Osadchuk, L.A.; Levina, N.B.; Telezhinskaya, I.N.; Khrapunov, S.N.; Berdyshev, G.D.; A  
Bioorg. Khim. 16, 448-455, 1990  
A>Title: Primary structure of main nuclear protein from headless mollusk Illex argentinus  
A:Reference number: PN0081; MUID:90329035  
A:Accession: PN0082  
A:Molecule type: protein  
A:Residues: 1-21 <OSA>  
A:Experimental source: sperm  
C:Superfamily: sperm histone  
C:Keywords: sperm

Query Match 24.3%; Score 35; DB 2; Length 21;  
Best Local Similarity 58.3%; Pred. No. 2.3e+02;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 KKRNR--RNSGCT 14  
| | | | |  
| | | | |  
Db 1 KKRNR--RNSGCT 12

RESULT 5  
JC4278  
ribosomal protein L41, cytosolic [similarity] - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 01-Sep-2000  
C:Accession: JC4278  
R:Chan, Y.L.; Olivera, J.; Wool, I.C.  
Biochem. Biophys. Res. Commun. 214, 810-818, 1995  
A>Title: The primary structures of rat ribosomal proteins L4 and L41.  
A:Reference number: JC4277; MUID:96024571

A:Accession: JC4278  
A:Molecule type: mRNA  
A:Residues: 1-25 <CHA>  
A:Cross-references: EMBL:X82550; NID:9575383; PIDN:CAA5789.1; PID:9575384  
C:Superfamily: rat ribosomal protein L41  
C:Keywords: ribosome

Query Match 24.3%; Score 35; DB 1; Length 25;  
Best Local Similarity 54.5%; Pred. No. 2.8e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 ERKKR--RNSGCT 12  
| | | | |  
| | | | |  
Db 14 KKRNR--RNSGCT 24

RESULT 6  
JQ1617  
ribosomal protein L41.e - human  
N:Alternate names: HG12 protein; ribosomal protein YL41  
C:Species: Homo sapiens (man)  
C>Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 18-Aug-2000  
C:Accession: JQ1617; JC5659; S31691  
R:Klaudy, J.; von der Kammer, H.; Scheit, K.H.  
Biochem. Biophys. Res. Commun. 187, 901-906, 1992  
A>Title: Characterization by cDNA cloning of the mRNA of a highly basic human protein  
A:Reference number: JQ1617; MUID:92412140  
A:Accession: JQ1617  
A:Molecule type: mRNA  
A:Residues: 1-25 <KLA>  
A:Cross-references: EMBL:Z12962; NID:936135; PIDN:CAA78306.1; PID:936136  
R:Lee, J.H.; Kim, J.M.; Kim, M.S.; Lee, Y.T.; Marshak, D.R.; Bae, Y.S.  
Biochem. Biophys. Res. Commun. 236, 462-467, 1997  
A>Title: The highly basic ribosomal protein L41 interacts with the beta subunit of pr  
A:Reference number: JC5659; MUID:97446005  
A:Accession: JC5659  
A:Molecule type: mRNA  
A:Residues: 1-25 <LEE>  
A:Cross-references: EMBL:Z12962; NID:936135; PIDN:CAA78306.1; PID:936136  
C:Comment: This protein stimulates phosphorylation of the beta chain of DNA topoisom  
C:Superfamily: rat ribosomal protein L41

Query Match 24.3%; Score 35; DB 2; Length 25;  
Best Local Similarity 54.5%; Pred. No. 2.8e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 ERKKR--RNSGCT 12  
| | | | |  
| | | | |  
Db 14 KKRNR--RNSGCT 24

RESULT 7  
JC4685  
ribosomal protein L41 - cat  
C:Species: Felis silvestris catus (domestic cat)  
C>Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #text\_change 18-Aug-2000  
C:Accession: JC4685  
R:Starkey, C.R.; Menon, R.P.; Prabh, S.; Levy, L.S.  
Biochem. Biophys. Res. Commun. 220, 648-652, 1996  
A>Title: Primary sequence and evolutionary conservation of ribosomal protein genes fr  
A:Reference number: JC4685; MUID:96183078  
A:Accession: JC4685  
A:Molecule type: mRNA  
A:Residues: 1-25 <STA>  
A:Cross-references: GB:U22229; NID:9950108; PIDN:AA801667.1; PID:9950109  
A:Experimental source: thymic lymphoma  
C:Genetics:  
A:Gene: rp1a1  
C:Superfamily: rat ribosomal protein L41

Query Match 24.3%; Score 35; DB 2; Length 25;  
Best Local Similarity 54.5%; Pred. No. 2.8e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 ERKRRORRRS 12  
:111:1:111  
Db 14 KRRKRRMORS 24

RESULT 8  
S38425  
ribosomal protein GlA1 - upland cotton  
C:Species: Gossypium hirsutum (upland cotton)  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 18-Aug-2000  
C:Accession: S38425  
R:Turley, R.B.; Ferguson, D.L.; Meredith, W.R.  
submitted to the EMBL Data Library, October 1993  
A:Reference number: S38425  
A:Accession: S38425  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-25 <TUR>  
A:Cross-references: EMBL:X75423; NID:g407800; PIDN:CA53175.1; PID:g407801  
C:Superfamily: rat ribosomal protein L41

Query Match 24.3%; Score 35; DB 2; Length 25;  
Best Local Similarity 54.5%; Pred. No. 2.8e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 ERKRRORRRS 12  
:111:1:111  
Db 14 KRRKRRMORS 24

RESULT 9  
T06233  
ribosomal protein l41 - soybean (fragment)  
C:Species: Glycine max (soybean)  
C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 18-Aug-2000  
C:Accession: T06233  
R:Mahalingam, R.; Knap, H.T.  
submitted to the EMBL Data Library, February 1998  
A:Reference number: Z15555  
A:Accession: T06233  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-25 <MAH>  
A:Cross-references: EMBL:AF047051; NID:g2905777; PIDN:AAC03557.1; PID:g2905778  
A:Experimental source: cultivar Essex; root  
C:Superfamily: rat ribosomal protein L41  
C:Keywords: protein biosynthesis; ribosome

Query Match 24.3%; Score 35; DB 2; Length 25;  
Best Local Similarity 54.5%; Pred. No. 2.8e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 ERKRRORRRS 12  
:111:1:111  
Db 14 KRRKRRMORS 24

RESULT 10  
T49214  
ribosomal protein GlA1-like - Arabidopsis thaliana  
N:Alternate names: protein F27K19.200  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 18-Aug-2000  
C:Accession: T49214  
R:Benes, V.; Wurmbech, E.; Drzonek, H.; Ansoorge, W.; Mewes, H.W.; Rudd, S.; Lemcke, K.;  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z25014

A:Accession: T49214  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-25 <BEN>  
A:Cross-references: EMBL:AL163832; GSPDB:GN00061; ATSP:F27K19.200  
A:Experimental source: cultivar Columbia; BAC clone F27K19  
C:Genetics:  
A:Gene: ATSP:F27K19.200  
A:Map position: 3  
A:Introns: 4/3  
C:Superfamily: rat ribosomal protein L41

Query Match 24.3%; Score 35; DB 2; Length 25;  
Best Local Similarity 54.5%; Pred. No. 2.8e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 ERKRRORRRS 12  
:111:1:111  
Db 14 KRRKRRMORS 24

RESULT 11  
SRAPC  
protamine B - Russian sturgeon  
N:Alternate names: sturine  
C:Species: Acipenser gueldenstaedti (Russian sturgeon)  
C>Date: 12-Aug-1981 #sequence\_revision 12-Aug-1981 #text\_change 16-Feb-1997  
C:Accession: A02665  
R:Yulikova, E.P.; Evseenko, L.K.; Baratova, L.A.; Belyanova, L.P.; Rybin, V.K.; Silae  
Bloorig, Khlm. 2, 1613-1617, 1976  
A:Reference number: A02665  
A:Accession: A02665  
A:Molecule type: protein  
A:Residues: 1-27 <YUL>  
C:Superfamily: protamine Y2  
C:Keywords: chromosomal protein; DNA binding; spermatogenesis

Query Match 23.6%; Score 34; DB 1; Length 27;  
Best Local Similarity 60.0%; Pred. No. 4e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 KRRRORRRSG 13  
:111:11111  
Db 11 QRKRRRRRSG 20

RESULT 12  
A38304  
heterogeneous ribonuclear particle protein A1 - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 14-Jun-1991 #sequence\_revision 14-Jun-1991 #text\_change 31-Oct-1997  
C:Accession: A38304  
R:Kumar, A.; Casas-Finet, J.R.; Lunneau, C.J.; Karpel, R.L.; Merrill, B.M.; Williams,  
J. Biol. Chem. 265, 17094-17100, 1990  
A:Title: Mammalian heterogeneous nuclear ribonucleoprotein A1. Nucleic acid binding p  
A:Reference number: A38304; MUID:91009136  
A:Accession: A38304  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-15 <KUM>  
C:Superfamily: helix-destabilizing protein; ribonucleoprotein repeat homology

Query Match 22.9%; Score 33; DB 2; Length 15;  
Best Local Similarity 87.5%; Pred. No. 3.2e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 QRKRRSG 15  
:1111111  
Db 8 QRKRRSG 15

## RESULT 13

T38719

ribosomal protein L41 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence-revision 03-Dec-1999 #text-change 18-Aug-2000

C:Accession: T38719; T38731; T39176

R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, February 1996

A:Reference number: 221807

A:Accession: T38731

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-25 &lt;NR&gt;

A:Cross-references: EMBL:269369; PIDN:CA940187.1; GSPDB:GN00066

A:Experimental source: strain 972h-; cosmid c3f10

R:Gentles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, September 1997

A:Reference number: 221797

A:Accession: T38731

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-25 &lt;GEN&gt;

A:Cross-references: EMBL:299167; PIDN:CA940152.1; GSPDB:GN00066; SPDB:SPAC366.13c

A:Experimental source: strain 972h-; cosmid c3c6

R:Bothe, G.; Pohl, T.; McDougall, R.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, August 1999

A:Reference number: 221833

A:Accession: T39176

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-25 &lt;BOT&gt;

A:Cross-references: EMBL:AL109738; PIDN:CA951162.1; GSPDB:GN00066; SPDB:SPAC8f11.01c

A:Experimental source: strain 972h-; cosmid c8f11

C:Genetics: 1

A:Gene: SPAC3f10.18c; SPAC366.13c; SPAC3f10.18c; SPAC8f11.01c

A:Map position: 1

A:Introns: 24/2

C:Superfamily: rat ribosomal protein L41

Query Match 22.9%; Score 33; DB 2; Length 25;  
 Best Local Similarity 54.5%; Pred. No. 5.1e+02;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

## RESULT 14

S35924

T-cell receptor gamma chain - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 06-Feb-1995 #sequence-revision 06-Feb-1995 #text-change 30-Jun-2001

C:Accession: S35924

R:Mathioudakis, G.; Platsoucas, C.D.

submitted to the EMBL Data Library, May 1993

A:Description: An alternative splicing between V-gamma<sub>2</sub> and C-gamma<sub>2</sub> gene seg

A:Reference number: S32764

A:Accession: S35924

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-29 &lt;MAT&gt;

A:Cross-references: EMBL:222684

C:Keywords: T-cell receptor

Query Match 22.2%; Score 32; DB 2; Length 29;  
 Best Local Similarity 43.8%; Pred. No. 8e+02;  
 Matches 7; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 6 RKQRRSGSGTDFLT 21

Db 3 RGMRLFGSGTLVVT 18  
 | : | | | | : |

## RESULT 15

C49042

Ig heavy chain V region, anti-DNA monoclonal G2a antibody - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 19-Dec-1993 #sequence-revision 18-Nov-1994 #text-change 16-Aug-1996

C:Accession: C49042

R:Takl, S.; Hirose, S.; Kinoshita, K.; Nishimura, H.; Shimamura, T.; Hamuro, J.; Shlr

Eur. J. Immunol. 22, 987-992, 1992

A:Title: Somatically mutated IgG anti-DNA antibody clonally related to germ-line enco

A:Reference number: A49042; MUID:92201320

A:Contents: NZB X NZW F1, hybridoma BW9-45

A:Accession: C49042

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-21 &lt;TAK&gt;

A:Note: sequence extracted from NCBI backbone (NCBIP:92267)

C:Keywords: heterotetramer; immunoglobulin

Query Match 21.5%; Score 31; DB 2; Length 21;  
 Best Local Similarity 66.7%; Pred. No. 8e+02;  
 Matches 8; Conservative 1; Mismatches 1; Indels 2; Gaps 1;

OY 13 GSGTDFLTGIS 24

Db 12 GGGT--TLTVSS 21

| : | | | |

Search completed: July 15, 2002, 13:52:41

Job time: 160 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:51:47 ; Search time 12.84 Seconds

(Without alignments)  
87.451 Million cell updates/sec

Title: US-09-712-819a-9

Sequence: 144  
1 YKKRRRRRRSGSGDTFTLTISLQAEQ 29

Scoring table: BLOSUM62  
Gapop 10.0 , Capext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 1766

Minimum DB seq length: 0  
Maximum DB seq length: 29

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	27.1	29	1 RL15_HALCU	P05971 halobacteri
2	38	26.4	27	1 PRTA_ACIST	P02324 acipenser s
3	35	24.3	25	1 RL41_ARATH	P35015 arabidopsis
4	35	24.3	25	1 RL41_HUMAN	P28751 homo sapien
5	35	24.3	25	1 RL41_SOYBN	O49224 glycine max
6	34	23.6	27	1 PRTB_ACIGU	P03323 acipenser g
7	33	22.9	25	1 RL41_SCHPO	O94710 schizosach
8	32	22.2	25	1 RL41_AGABI	P78569 agarticus bl
9	32	22.2	25	1 RL41_QUEST	O82713 quercus sub
10	31	21.5	25	1 RL41_YEAST	P05746 saccharomyc
11	29	20.1	22	1 RL41_METJA	P54025 methanococc
12	28	19.4	27	1 HM12_CABEL	P17487 caenorhabdi
13	28	19.4	29	1 TL16_SPIOI	P81834 spinacia ol
14	27.5	19.1	29	1 GLUC_LAMFL	O94939 lampetra fl
15	27	18.8	13	1 NEUR_CAVPO	P33560 cavia porce
16	26	18.1	20	1 PSBW_WHEAT	P55967 triticum ae
17	25	17.4	22	1 RL18_HALME	P50561 halobacteri
18	25	17.4	23	1 RL18_HALVO	P50563 halobacteri
19	25	17.4	24	1 HEMO_LINRE	P23544 lingua ree
20	25	17.4	25	1 SCRK_LACLA	O09124 lactococcus
21	25	17.4	26	1 RL18_HALLA	P05606 halobacteri
22	24	16.7	29	1 GLUC_DIDMA	P13108 didelphis m
23	24	16.7	29	1 GLUC_RABIT	P24408 onycholagus
24	23	16.0	15	1 UC06_MAIZE	P08612 zea mays (m
25	23	16.0	24	1 VGI_BPAL3	P08766 bacterioph
26	23	16.0	25	1 RS11_THRTH	P08376 thermus aqu
27	23	16.0	29	1 VGI_SPPV	P11341 spiroplasma
28	22	15.3	12	1 GLUC_CALMI	P11389 callorhynch
29	22	15.3	11	1 CA41_LITCI	P82091 littoria cit
30	22	15.3	12	1 UR2A_CATCO	P04558 catostomus
31	22	15.3	20	1 TENB_ACTTE	P30634 actinia ten
32	22	15.3	23	1 PQ0A_KLEPN	P27503 klebsiella
33	22	15.3	29	1 RP54_CLOXL	P38944 clostridium

34	21	14.6	9	1 D1_NEPNO	P24816 nephrops no
35	21	14.6	9	1 FAR6_CALVO	P41861 calliphora
36	21	14.6	11	1 CA42_LITCI	P82092 littoria cit
37	21	14.6	16	1 RIKK_TIRIKI	P16093 trichosan
38	21	14.6	18	1 YAS5_RHOPA	O02005 rhodosphe
39	21	14.6	19	1 LANA_ACTTG	P36650 actinoplane
40	21	14.6	20	1 TENA_ACTTE	P30833 actinia ten
41	21	14.6	22	1 IFZG_PIG	P20461 sus scrofa
42	21	14.6	22	1 MOTI_CANFA	P19863 canis fami
43	21	14.6	22	1 TL11_SPIOI	P82657 spinacia ol
44	21	14.6	25	1 ANDT_ANDAN	P56684 androcione
45	21	14.6	25	1 VGI_BP04	P03652 bacterioph

## ALIGNMENTS

RESULT 1					
ID	RL15_HALCU	STANDARD;	PRT;	29 AA.	
AC	P05971				
DT	01-NOV-1988 (Rel. 09, Created)				
DT	01-NOV-1988 (Rel. 09, Last sequence update)				
DE	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	50S ribosomal protein L15P (H116) (Fragment).				
GN	RPL15P.				
OS	Halobacterium cutirubrum.				
OC	Archaea; Buryarchaeota; Halobacteriales; Halobacteriaceae;				
OC	Halobacterium.				
OX	NCBI_TaxID=2242;				
RN	[1]				
RP	SEQUENCE.				
RA	MEDLINE=84282108; PubMed=6467081;				
RA	Matheson A.T., Yaguchi M., Christensen P., Rollin C.F., Hasnain S.;				
RT	"Purification, properties, and N-terminal amino acid sequence of				
RT	certain 50S ribosomal subunit proteins from the archaebacterium				
RT	Halobacterium cutirubrum.";				
RL	Can. J. Biochem. Cell Biol. 62:426-433(1984).				
CC	-1 SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.				
DR	PIR; S08555; S08555.				
DR	InterPro; IPR001196; Ribosomal_L15.				
DR	PROSITE; PS00475; RIBOSOMAL_L15; PARTIAL.				
KW	Ribosomal protein.				
FT	NON_TER 29				
FT	SEQUENCE 29 AA; 3170 MW; 73D130831B4BE790 CRC64;				
Query Match 27.1%; Score 39; DB 1; Length 29;					
Best Local Similarity 66.7%; Pred. No. 12;					
Matches 10; Conservative 0; Mismatches 3; Indels 2; Gaps 1;					
OY	4 KRRRR--RRSGGT 16				
DB	3 KRRRRGSRTHGGT 17				
RESULT 2					
ID	PRTA_ACIST	STANDARD;	PRT;	27 AA.	
AC	P02324;				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	01-NOV-1995 (Rel. 32, Last annotation update)				
DE	Proteomines A and C (Stellin A and C).				
OS	Acipenser stellatus (Sevruga).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Actinopterygii; Chondrostei; Acipenseriformes; Acipenseridae;				
OC	Acipenserinae; Acipenser.				
OX	NCBI_TaxID=7903;				
RN	[1]				
RP	SEQUENCE (STELLIN A).				
RA	Yulikova E.P., Rybin V.K., Silaev A.B.;				
RA	"The primary structure of stellin A.";				
RT					

RL Bloorg. Rhm. 5:5-10(1979).  
RN [2]  
RP SEQUENCE (STELLIN C).  
RX MEDLINE-94019236; PubMed-1339950;  
RA Rylm V.K., Revina L.P., Baratova L.A.;  
RT "Comparison of amino acid sequences of sturgeon triptamines using  
RT proteinases from *Aplocheilichthys stellatus* gonads as an example.";  
RL Mol. Biol. (Mosk) 26:300-306(1992).  
CC -1- FUNCTION: PROTEINASE SUBSTITUTION FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: TESTIS.  
DR PIR: A02666; SRAPAS.  
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
KW Testis; DNA condensation; Nucleolar protein.  
FT VARIANT 1 1 MISSING (IN STELLIN C).  
SQ SEQUENCE 27 AA; 3532 MW; 3EAA5B950FA8658 CRC64;  
  
OY 4 KKRORRRSG 13 26.4%; Score 38; DB 1; Length 27;  
DB 13 KRRRRRRRHG 22 70.0%; Pred. No. 16;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
RESULT 3  
RL41\_ARATH STANDARD; PRT: 25 AA.  
ID RL41\_ARATH  
AC P35015;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DE 60S ribosomal protein L41.  
GN (RPL41-A OR ATG56045 OR T6H22.15) AND (RPL41-B OR ATG560520 OR  
GN T8624.5 OR T8624.5) AND (RPL41-C OR ATG311120 OR F9F8.7) AND  
GN (RPL41-D OR ATG356020 OR F27K19-200).  
OS Arabidopsis thaliana (Mouse-ear cress).  
OS Nicotiana tabacum (Common tobacco).  
OS Gossypium hirsutum (Upland cotton).  
OS Pisum sativum (Garden pea).  
OS Hordeum vulgare (Barley), and  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eustoids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702, 4097, 3635, 3888, 4513, 4530;  
RN [1]  
RP SEQUENCE FROM N.A. (ATG56045).  
RC SPECIES-A.thaliana; STRAIN-CV. COLUMBIA;  
RX MEDLINE-21016719; PubMed-1130712.  
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
RA Dunn P., Elgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huzar L.,  
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kvan A., Lam B.,  
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,  
RA Miltiescher J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I.,  
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
RA Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D.,  
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
RT thaliana.";  
RL Nature 408:816-820(2000).  
  
RN [2]  
RP SEQUENCE FROM N.A. (ATG311120 AND ATG56020).  
RC SPECIES-A.thaliana; STRAIN-CV. COLUMBIA;  
RX MEDLINE-21016720; PubMed-1130713;  
RA Salanoubat M., Lemck C., Rieger M., Ansoerge W., Unseld M.,  
RA Fattman B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,  
RA Delserny M., Boutry M., Givelli L.A., Mache R., Pildgomech P.,  
RA De Simone V., Cholsine N., Artiguenave F., Robert C., Brolier P.,  
RA Mincker P., Catolico L., Weissenbach J., Saurin W., Queller F.,  
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
RA Humbach E., Dizonex H., Effle H., Jordan N., Baiget S.,  
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,  
RA Verzi A., D'Angelo M., Pallavicini A., Toppo S., Stindl B.,  
RA Conrad A., Hornischer K., Kauer G., Lochner T.-H., Nordstiek G.,  
RA Reichelt J., Scharte M., Schoen O., Bagues M., Terol J., Clement J.,  
RA Navarro P., Collado C., Perez-Perez A., Oetgenwelder B., Duchemin D.,  
RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Maury D.,  
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casachueria E.,  
RA Monfort A., Argireu A., Flores M., Ligouri R., Vitale D.,  
RA Manhaupf G., Haase D., Schoof H., Rued S., Zaccaria P., Mexes H.-M.,  
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
RA Rooney T., Rizzo M., Maiti A., Uterback T., Fujii C.Y., Shea T.P.,  
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,  
RA Pal G., Miltiescher J., Sellers P., Gill J.E., Feldblyum T.V.,  
RA Preuss D., Lin X., Niernan W.C., Salzberg S.L., White O., Venter J.C.,  
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,  
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,  
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,  
RA Nakanebe A., Yamada M., Yasuda M., Tabata S.;  
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
RT thaliana.";  
RL Nature 408:820-822(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES-A.thaliana; STRAIN-CV. COLUMBIA;  
RX Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
RA Bowser L., Carrincci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
RA Ishida J., Jiang P.X., Jones T., Kamya A., Karlin-Neumann G.,  
RA Kawai J., Lam B., Lee J.M., Lin X., Liu S.X., Miranda M., Natuska M.,  
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
RA Satou M., Seki M., Southwick A., Tang C.C., Tortum M., Yamada K.,  
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
RA Ecker J.R.;  
RT "Arabidopsis cDNA clones";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES-N.tabacum; STRAIN-CV. SRL; TISSUE=Leaf;  
RX Zhou X.R.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC SPECIES-G.hirsutum; STRAIN-CV. DELTA PINE 62;  
RX MEDLINE-93062728; PubMed-7972506;  
RA Turley R.B., Ferguson D.L., Meredith M.R.;  
RT "Isolation and characterization of a cDNA encoding ribosomal protein  
RT L41 from cotton (*Gossypium hirsutum* L.).";  
RL Plant Physiol. 105:1449-1450(1994).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC SPECIES-P.sativum; STRAIN-CV. LITTLE MARVEL; TISSUE=Root tip;  
RX Woo H.H.;  
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP SEQUENCE FROM N.A.  
RC SPECIES-H.vulgare; STRAIN-CV. BOMI; TISSUE=Endosperm;  
RX Resmussen S.K.;  
RT "Barley L41 ribosomal protein from immature endosperm.";  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RP SEQUENCE FROM N.A.  
RC SPECIES-O.sativa; STRAIN-CV. Nippondare;

RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,  
 RA Zisman V., Pal G., Bowman C.L., Fujii C.Y., VanAken S.E.,  
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,  
 RA Quakenbush J., White O., Salzberg S.L., Fraser C.M.,  
 RT "Oriza sativa chromosome 3 BAC OSJNBa0091j19 genomic sequence.",  
 CC Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RL  
 CC -1- SIMILARITY: BELONGS TO THE LALE FAMILY OF RIBOSOMAL PROTEINS.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL: AC009894; AAF02845.1; -  
 DR EMBL: AC074395; AAS0829.1; -  
 DR EMBL: AC009991; AAF01511.1; -  
 DR EMBL: AL163832; CAB87856.1; -  
 DR EMBL: AV058053; AAL24161.1; -  
 DR EMBL: U26255; AAA67297.1; -  
 DR EMBL: X75423; CAA53175.1; -  
 DR EMBL: L47967; AAF79268.1; -  
 DR EMBL: AJ001160; CAA04564.1; -  
 DR EMBL: AC084320; AAK09215.1; -  
 DR PIR: S38425; S38425.  
 KW Ribosomal protein.  
 SQ SEQUENCE 25 AA; 3428 MW; 9AD629D4293C039E CRC64;  
 -----  
 QY 2 ERKKRORRRS 12  
 :11:1:1:11  
 DB 14 KRKKRKRORS 24  
 -----  
 RESULT 4  
 ID RL41\_HUMAN STANDARD; PRT; 25 AA.  
 AC P28751;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last annotation update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE 60S ribosomal protein L41 (Hs12).  
 GN RPL41.  
 OS Homo sapiens (Human),  
 OS Mus musculus (Mouse),  
 OS Rattus norvegicus (Rat),  
 OS Felis silvestris catus (Cat), and  
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
 CC Fukuyama, Metazoa: Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC NCBI\_TaxID=9606, 10090, 10116, 9683, 31033;  
 OX [1]  
 RN RP  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Human; TISSUE-Ovary;  
 RX MEDLINE-92412140; PubMed-1326959;  
 RA Klaudiny J., von der Kammer H., Scheit K.H.;  
 RT "Characterization by cDNA cloning of the mRNA of a highly basic human  
 RT protein homologous to the yeast ribosomal protein Y141.";  
 RL Biochem. Biophys. Res. Commun. 187:901-906(1992).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Human;  
 RX MEDLINE-97446005; PubMed-9299532;  
 RA Lee J.-H., Kim J.-M., Kim M.-S., Lee Y.-T., Marshak D.R., Bae Y.-S.;  
 RT "The highly basic ribosomal protein L41 interacts with the beta  
 RT subunit of protein kinase CKII and stimulates phosphorylation of DNA  
 RT topoisomerase IIalpha by CKII.";

RL Biochem. Biophys. Res. Commun. 238:462-467(1997).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RC SPECIES-Human;  
 RX Go H., Miyado K., Hasuwa H., Taniguchi S.;  
 RT "Characterization of human ribosomal protein L41 genomic structure.";  
 RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 RL [4]  
 RN RP  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Human;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babage A.K., Baguley C.L.,  
 RA Bailey O.P., Barlow K.P., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Copley V.E., Collier R.E., Connor R., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Levaslainho M.H., Leverisa M., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConachie L.J., McIay K., McMurtry A.A.,  
 RA Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,  
 RA Skuse C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaubin M., Walli M., Wallis J.M., Williams S.A.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams D.R., Beck S.,  
 RA Wilming L., Wray P.W., Hubbard T., Dudbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RT Nature 414:865-871(2001).  
 RL [5]  
 RN RP  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Rat; STRAIN-SPRAGUE-DAWLEY; TISSUE-Liver;  
 RA Chan Y.-L., Wool I.G.;  
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
 RL [6]  
 RN RP  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Mouse; STRAIN-C57BL/6;  
 RA Rocha D., Anderson E., Botcherby M., Jordan B.;  
 RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.  
 RL [7]  
 RN RP  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-F.s.catus;  
 RX MEDLINE-96183078; PubMed-8607819;  
 RA Starkey C.R., Menon R.P., Prabhu S., Levy L.S.;  
 RT "Primary sequence and evolutionary conservation of ribosomal protein  
 RT genes from the domestic cat.";  
 RL Biochem. Biophys. Res. Commun. 220:648-652(1996).  
 RN [8]  
 RN RP  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-F.rubripes;  
 RX MEDLINE-9917347; PubMed-10077531;  
 RA Geilner K., Brenner S.;  
 RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu  
 RT rubripes.";  
 RL Genome Res. 9:251-258(1999).  
 CC -1- FUNCTION: INTERACTS WITH THE BETA SUBUNIT OF PROTEIN KINASE CKII  
 CC AND STIMULATES PHOSPHORYLATION OF DNA TOPOISOMERASE II ALPHA BY  
 CC CKII.  
 CC -1- SIMILARITY: BELONGS TO THE LALE FAMILY OF RIBOSOMAL PROTEINS.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

```

CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: Z12962; CAA78306.1; -
DR  EMBL: AF026844; AAB82715.1; -
DR  EMBL: AB010874; BAA31508.1; -
DR  EMBL: AL035562; CAA46994.1; -
DR  EMBL: X82550; CAA57899.1; -
DR  EMBL: U93862; AAB52254.1; -
DR  EMBL: U22229; AAB01657.1; -
DR  EMBL: AF056116; AAC34393.1; -
DR  PIR: J01617; J01617.
DR  PIR: S31691; S31691.
KW  Ribosomal protein.
SQ  SEQUENCE 25 AA: 3456 MW: 9AD629C69ECC039E CRC64;

Query Match 24.3%: Score 35; DB 1; Length 25;
Best Local Similarity 54.5%: Pred. No. 39;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 ERKRRRRRS 12
    :|:|:|:|:|:|
DB 14 KRRRRRRRS 24

RESULT 5
ID RL41_SOYBN STANDARD: PRT: 25 AA.
AC 049224;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L41.
GN RPL41.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
OC eurosid1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ESSEX; TISSUE=Root;
RA WOO H.-H.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L41E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: AF047051; AAC03557.1; -
KW  Ribosomal protein.
SQ  SEQUENCE 25 AA: 3130 MW: AD933FB8E49EDD6E CRC64;

Query Match 24.3%: Score 35; DB 1; Length 25;
Best Local Similarity 54.5%: Pred. No. 39;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 ERKRRRRRS 12
    :|:|:|:|:|:|
DB 14 KRRRRRRRS 24

RESULT 6
ID PRTB_ACI0U STANDARD: PRT: 27 AA.
AC P02323;

```

```

DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Protamine B (Sturine B) (Stellin B).
OS Acipenser guldenstadti (Caspian sturgeon) (Russian sturgeon), and
OC Acipenser stellatus (Sevruga).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Chondrostei; Acipenseriformes; Acipenseridae;
OC Acipenserinae; Acipenser.
OX NCBI_TaxID=7902; 7903;
RN [1]
RP SEQUENCE.
RC SPECIES=A.guldenstadti;
RA Yulikova E.P., Evseenko L.K., Baratova L.A., Bolyanova L.P.,
RA Rybin V.K., Silaev A.B.;
RT "The primary structure of sturine B, a protamine from Caspian
RT sturgeon."
RT Bioorg. Khim. 2:1613-1617(1976).
RN [2]
RP SEQUENCE.
RC SPECIES=A.stellatus;
RA Rybin V.K., Yulikova E.P.;
RA Khim. Prirod. Soedin. 5:700-704(1979).
CC -1- FUNCTION: PROTAGINIS SUBSTITUTE FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: TESTIS.
DR PIR: A02665; SRAP.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
SQ SEQUENCE 27 AA: 3707 MW: E300B4B1737C80 CRC64;

Query Match 23.6%: Score 34; DB 1; Length 27;
Best Local Similarity 60.0%: Pred. No. 60;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 KRRRRRRSG 13
    :|:|:|:|:|
DB 11 QRRRRRRRG 20

RESULT 7
ID RL41_SCHPO STANDARD: PRT: 25 AA.
AC Q9Y710;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60S ribosomal protein L41.
GN (RPL41A OR SPAC36.13C) AND (RPL41B OR SPAC31F10.18C OR SPAC8F11.01C).
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A. (RPL41A).
RC STRAIN=972;
RA Genies S., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (RPL41B).
RC STRAIN=972;
RA Bothe G., Pohl T., McDougall R., Rajandream M.A., Barrell B.G.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L41 IN S.POMBE.
CC -1- SIMILARITY: BELONGS TO THE L41E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its

```



```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch.)
CC -----
DR EMBL: Z99167; CAB40152.1; -
DR EMBL: Z69369; CAB40187.1; -
DR EMBL: AL109738; CAB52162.1; -
KW Ribosomal protein; Multigene family.
SQ SEQUENCE 25 AA; 3411 MW; 96C629DD9ECC0535 CRC64;

Query Match      22.9%; Score 33; DB 1; Length 25;
Best Local Similarity 34.5%; Pred. No. 78;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 ERKKRRRRRS 12
   :|:|:|:|:|
Db 14 KRRRRRRARS 24

RESULT 8
RL1_AGABI
ID RL1_AGABI STANDARD; PRT; 25 AA.
AC P78569;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 60S ribosomal protein L41.
GN RL41.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HORST U1;
RA Sonnenberg A.S.M., de Groot P.W.J., Schaap P.J., Baars J.J.P.,
RA Visser J., van Gilsenven L.J.L.D.;
RT "Isolation of expressed sequence tags of Agaricus bisporus and their
RT assignment to chromosomes."
RL Appl. Environ. Microbiol. 62:4542-4547(1996).
CC -!- SIMILARITY: BELONGS TO THE L41E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch.)
CC -----
DR EMBL: X94764; CA64390.1; -
KW Ribosomal protein.
SQ SEQUENCE 25 AA; 3413 MW; A97629DD0143C04EA CRC64;

Query Match      22.2%; Score 32; DB 1; Length 25;
Best Local Similarity 60.0%; Pred. No. 1,1e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 RKKRRRRRS 12
   :|:|:|:|
Db 15 RRRRRRRARS 24

RESULT 9
RL1_QUEST
ID RL1_QUEST STANDARD; PRT; 25 AA.
AC 082713;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

```

```

DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L41.
GN RPL41.
OS Quercus suber (Cork oak).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fagales; Fagaceae; Quercus.
OX NCBI_TaxID=58331;
RN [1]
RP SEQUENCE FROM N.A.
RA Hugnet G., Pla M., Verdaguer D., Molinas M.;
RT "Ribosomal proteins in Quercus suber."
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE L41E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch.)
CC -----
DR EMBL: AJ001347; CA404691.1; -
KW Ribosomal protein.
SQ SEQUENCE 25 AA; 3337 MW; 6BC629DD9EDB3383 CRC64;

Query Match      22.2%; Score 32; DB 1; Length 25;
Best Local Similarity 54.5%; Pred. No. 1,1e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 ERKKRRRRRS 12
   :|:|:|:|
Db 14 KRRRRRRARS 24

RESULT 10
RL1_YEAST
ID RL1_YEAST STANDARD; PRT; 25 AA.
AC P05746;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L41 (YL41) (YL47).
GN (RPL41A OR RPL47A OR YL41A OR YDL184C OR D1290) AND
GN (RPL41B OR RPL47B OR YL41B OR YDL133BC).
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A. (RPL41A AND RPL41B).
RX MEDLINE=90254826; PubMed=2187623;
RA Suzuki K., Hashimoto T., Otsuka E.;
RT "Yeast ribosomal proteins: XI. Molecular analysis of two genes
RT encoding YL41, an extremely small and basic ribosomal protein, from
RT Saccharomyces cerevisiae."
RL Curr. Genet. 17:185-190(1990).
RN [2]
RP SEQUENCE FROM N.A. (RPL41A).
RX STRAIN=S288C / FY1679;
RC MEDLINE=96021607; PubMed=8533471;
RA Vernasselt P., Voet M., Volckaert G.;
RT "New open reading frames, one of which is similar to the n1y gene of
RT of Saccharomyces cerevisiae."
RL Yeast 11:961-966(1995).
RN [3]
RP SEQUENCE FROM N.A. (RPL41B).
RX STRAIN=S288C / FY1679;
RC MEDLINE=97127826; PubMed=8972577;
RA Woeifl S., Haneman V., Saluz H.P.;

```

```

RT      "Analysis of a 26,756 bp segment from the left arm of yeast
RL      chromosome IV."
RM      Yeast 12:1549-1554(1996).
RA      SEQUENCE.
RA      Olaka E., Higo K.-I., Itoh T.
RT      "Yeast ribosomal proteins: VII. Isolation of two proteins and
RT      sequence characterization of twenty-four proteins from cytoplasmic
RT      ribosomes."
RL      Mol. Gen. Genet. 195:544-546(1984).
CC      -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L41 IN YEAST.
CC      -1- SIMILARITY: BELONGS TO THE L41E FAMILY OF RIBOSOMAL PROTEINS.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL; X16065; CAA34201.1; -
DR      EMBL; X16066; CAA34202.1; -
DR      EMBL; X83276; CAA58262.1; -
DR      EMBL; Z74232; CAA98795.1; -
DR      EMBL; Z74181; CAA98705.1; -
DR      EMBL; Z74182; CAA98706.1; -
DR      EMBL; X96876; CAA65626.1; -
DR      PIR; S22247; R6BYA4.
DR      PIR; S22346; R6BYA4.
DR      SGD; S0002343; RPL41A.
DR      SGD; S0002293; RPL41B.
RM      Ribosomal protein; Multigene family.
KW      Ribosome
SO      SEQUENCE 25 AA; 3337 MW; BD2629DD9ED85381 CRC64;

Query Match          21.5%  Score 31; DB 1; Length 25;
Best Local Similarity 54.5%  Pred. No. 1.5e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY      2 ERKRRRRRRS 12
DB      14 KRKRKRRARS 24

RESULT 11
RL41_METJA
ID      RL41_METJA          STANDARD;          PRT;          22 AA.
AC      P54025.
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      50S ribosomal protein L41E.
GN      RPL41E OR M20242.
OS      Methanococcus jannaschii.
OC      Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC      Methanococcus.
CX      NCBI_TaxID=2190;
RM      (1)
RM      SEQUENCE FROM N.A.
RC      STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX      MEDLINE=96337999; PubMed=8688087;
RA      Bolt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA      Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA      Kertavagis A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA      Overbeek R., Kirkness E.F., Weinstock K.G., Merriam J.M., Glodek A.,
RA      Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhlmann J.L., Nguyen D.,
RA      Uteback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA      Cotton M.D., Roberts C.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA      Klenk H.-P., Frazer C.M., Smith H.O., Woese C.R., Venter J.C.;
RT      "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT      jannaschii."
RL      Science 273:1058-1073(1996).

```

```

CC      -1- SIMILARITY: BELONGS TO THE L41E FAMILY OF RIBOSOMAL PROTEINS.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL; U67480; AAB98230.1; -
DR      TIGR; M20242; -
RM      Ribosomal protein; Complete proteome.
KW      Ribosome
SO      SEQUENCE 22 AA; 3089 MW; F948D74DD4EC98CB CRC64;

Query Match          20.1%  Score 29; DB 1; Length 22;
Best Local Similarity 41.7%  Pred. No. 2.6e+02;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY      3 RKRKRORRRSGS 14
DB      11 RLKRKRKRARS 22

RESULT 12
HM12_CAEEL
ID      HM12_CAEEL          STANDARD;          PRT;          27 AA.
AC      P17487.
DT      01-AUG-1990 (Rel. 15, Created)
DT      01-AUG-1990 (Rel. 15, Last sequence update)
DT      15-JUL-1998 (Rel. 36, Last annotation update)
DE      Homeobox protein cch-12 (Fragment).
GN      CCH-12.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC      Rhabditidae; Pelodierinae; Caenorhabditis.
CX      NCBI_TaxID=6239;
RM      (1)
RM      SEQUENCE FROM N.A.
RX      MEDLINE=90245646; PubMed=1970877;
RA      Schaller D., Wittmann C., Spicher A., Mueller F., Tobler H.;
RT      "Cloning and analysis of three new homeobox genes from the nematode
RT      Caenorhabditis elegans."
RL      Nucleic Acids Res. 18:2033-2036(1990).
CC      -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL; X17076; CAA34928.1; -
DR      PIR; S09504; S09504.
DR      HSSP; P02836; 1ENH.
DR      TRANSFAC; T02987; -
DR      InterPro; IPR001356; Homeobox.
DR      Pfam; PF00046; homeobox_1.
DR      PROSITE; PS00027; HOMEBOX_1; PARTIAL.
DR      PROSITE; PS00071; HOMEBOX_2; 1.
KW      Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT      DNA_BIND 1
FT      NON_TER 1
FT      DNA_BIND 16
FT      NON_TER 27
FT      NON_TER 27
SO      SEQUENCE 27 AA; 3182 MW; FDAFSD16F3870A5 CRC64;

Query Match          19.4%  Score 28; DB 1; Length 27;
Best Local Similarity 31.8%  Pred. No. 4.6e+02;
Matches 7; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:52:12 ; Search time 39.48 Seconds

(Without alignments)  
127.073 Million cell updates/sec

Title: US-09-712-819a-9

Perfect score: 144  
Sequence: 1 YERKKRRRRSGSGTDTLTITSSLDAD 29

Scoring table: BLOSUM62  
Gapop 10.0 , Capext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 12865

Minimum DB seq length: 0  
Maximum DB seq length: 29

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP:archaea:\*\n2: SP:bacteria:\*\n3: SP:fungi:\*\n4: SP:human:\*\n5: SP:invertebrate:\*\n6: SP:mammal:\*\n7: SP:mhc:\*\n8: SP:organelle:\*\n9: SP:phage:\*\n10: SP:plant:\*\n11: SP:rodent:\*\n12: SP:virus:\*\n13: SP:vertebrate:\*\n14: SP:unclassified:\*\n15: SP:virus:\*\n16: SP:bacteriap:\*\n17: SP:archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	43	29.9	26	13	Q90Y47 cyanoysilta
2	43	29.9	26	13	Q90VV3 ara maracan
3	43	29.9	26	13	Q90VV2 cyanoysilta
4	42	29.2	23	8	Q35223 oenothera b
5	37	25.7	13	12	Q65331 autographa
6	35	24.3	25	13	Q90VS7 ictalurus p
7	33	22.9	25	5	Q96282 spodoptera
8	33	22.9	27	13	Q57542 lampetra pl
9	33	22.9	27	13	Q57543 lampetra pl
10	32	22.2	29	6	Q9RS57 felis silve
11	31	21.5	23	15	Q85496 bovine leuk
12	30	20.8	23	5	Q94781 trypanosoma
13	30	20.8	28	10	Q42364 triticum ae
14	29	20.1	22	13	Q919P3 ara militar
15	29	20.1	28	7	Q19358 bos taurus
16	29	20.1	29	15	Q91Q15 human immun

17	28	19.4	20	12	Q85617 reovirus (t
18	28	19.4	21	12	Q9JG30 tt virus. o
19	28	19.4	21	12	Q9JG26 tt virus. o
20	28	19.4	21	12	Q9JG23 tt virus. o
21	28	19.4	22	12	Q85664 reovirus (t
22	28	19.4	23	8	Q957T0
23	28	19.4	23	8	Q957S8
24	28	19.4	26	11	Q99M13
25	28	19.4	27	5	Q95X71
26	28	19.4	28	12	Q9NNI8
27	28	19.4	28	12	Q9NNI0
28	28	19.4	29	15	P89816
29	28	19.4	29	15	P89821
30	27	18.8	15	13	Q9PRM3
31	27	18.8	22	2	Q9R3F6
32	27	18.8	22	8	Q957T4
33	27	18.8	24	11	Q61946
34	27	18.8	25	4	Q96PE8
35	27	18.8	26	2	Q9RA58
36	27	18.8	26	8	P92546
37	27	18.8	26	15	Q04070
38	27	18.8	27	12	Q9QHB7
39	27	18.8	28	12	Q9NNI9
40	26.5	18.4	15	2	Q47612
41	26.5	18.4	29	4	Q9UN87
42	26.5	18.4	29	15	Q91Q16
43	26	18.1	9	6	Q9T777
44	26	18.1	9	10	Q940K4
45	26	18.1	13	7	Q19690

#### ALIGNMENTS

RESULT 1					
Q90Y47					
ID Q90Y47:	PRELIMINARY:	PRT:	26 AA.		
DT 01-DEC-2001 (TRENBLREL. 19, Created)					
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)					
DE 01-DEC-2001 (TRENBLREL. 19, Last annotation update)					
DE CHROMO-HELICASE DNA BINDING PROTEIN-1W (FRAGMENT).					
CHD1-W.					
OS Cyanoysilta spixili x Ara maracana.					
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
OC Archosauria; Aves; Neognathae; Psittaciformes; Psittacidae.					
OX NCBI_TaxID=157078;					
RN [1]					
RP SEQUENCE FROM N.A.					
RA Miyaki C.Y., Farra P.J., Griffiths R., Araujo J.C.C., Barros Y.M.,					
RT "The last Spix's Macaw and an Illiger's Macaw produced a hybrid.*"					
RL Conserv. Genet. 2:53-55(2001).					
DR EMBL: AF276769; AAK96044.1; -.					
KW Helicase.					
FT NON_TER	1	26			
FT NON_TER	1	26			
SQ SEQUENCE	26 AA:	2973 MW:	F179E3EB84AC1297 CMC64:		
Query Match	29.9%	Score 43:	DB 13:	Length 26:	
Best Local Similarity	56.2%	Pred. No. 10:			
Matches 9;	Conservative 2;	Mismatches 5;	Indels 0;	Gaps 0;	
QY 2 ERKKRRRRSGSGTD 17					
DB 7 ERNRHSRRYSGSDSD 22					
RESULT 2					
Q90VV3					
ID Q90VV3:	PRELIMINARY:	PRT:	26 AA.		
DT 01-DEC-2001 (TRENBLREL. 19, Created)					

DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 GN CHROMO-HELICASE DNA BINDING PROTEIN-1W (FRAGMENT).  
 OS CHD1-W.  
 OC Aca. maracana.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Psittaciformes; Psittacidae; Ara.  
 OX NCBI\_TaxID=157077;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1, AND 2;  
 RA Miyaki C.Y., Farla P.J., Griffiths R., Araujo J.C.C., Barros Y.M.;  
 RL "The last Spix's Macaw and an Illiger's Macaw produced a hybrid";  
 RT Conserv. Genet. 2:53-55(2001).  
 DR EMBL: AF276764; AAK96039.1; -;  
 DR EMBL: AF276766; AAK96041.1; -;  
 KW Helicase.  
 FT NON\_TER  
 SQ SEQUENCE 26 AA; 2973 MW; F179E3EB84AC1297 CRC64;

Query Match 29.9%; Score 43; DB 13; Length 26;  
 Best Local Similarity 56.2%; Pred. No. 10;  
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 ERKKRRRRSGSGT 17  
 DB 7 ERHRSRRYSGSDD 22

RESULT 3  
 ID 090VV2 PRELIMINARY; PRT; 26 AA.  
 AC 090VV2:  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 GN CHROMO-HELICASE DNA BINDING PROTEIN-1W (FRAGMENT).  
 OS CHD1-W.  
 OC Cyanopectita spixii.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Psittaciformes; Psittacidae;  
 OC Cyanopectita.  
 OX NCBI\_TaxID=51902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1, AND 2;  
 RA Miyaki C.Y., Farla P.J., Griffiths R., Araujo J.C.C., Barros Y.M.;  
 RL "The last Spix's Macaw and an Illiger's Macaw produced a hybrid";  
 RT Conserv. Genet. 2:53-55(2001).  
 DR EMBL: AF276767; AAK96042.1; -;  
 DR EMBL: AF276768; AAK96043.1; -;  
 KW Helicase.  
 FT NON\_TER  
 SQ SEQUENCE 26 AA; 2973 MW; F179E3EB84AC1297 CRC64;

Query Match 29.9%; Score 43; DB 13; Length 26;  
 Best Local Similarity 56.2%; Pred. No. 10;  
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 ERKKRRRRSGSGT 17  
 DB 7 ERHRSRRYSGSDD 22

RESULT 4  
 ID 035223 PRELIMINARY; PRT; 23 AA.  
 AC 035223:  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 GN RIBOSOMAL PROTEIN 519 (FRAGMENT).  
 OS RPS19.  
 OC Oenothera biitiana (Bertero's evening primrose).  
 OC Mitochondrion.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Myrtales; Onagraceae; Oenothera.  
 OX NCBI\_TaxID=3950.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94250844; PubMed=8193306;  
 RA Beck H., Brennicke A., Schuster W.;  
 RL "Rps3 and rpl16 genes do not overlap in Oenothera mitochondria: GTC  
 as a potential translation initiation codon in plant mitochondria?";  
 RL Plant Mol. Biol. 24:811-818(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92107682;  
 RA Schuster W., Brennicke A.;  
 RL "RNA editing makes mistakes in plant mitochondria: editing looses  
 RT sense in transcripts of a rps19 pseudogene and in creating stop codons  
 RT in cox1 and rps3 mRNAs of Oenothera";  
 RL Nucleic Acids Res. 19:6923-6928(1991).  
 DR EMBL: X69140; CAA48933.1; -;  
 KW Mitochondrion.  
 FT NON\_TER  
 SQ SEQUENCE 23 AA; 2764 MW; 08B5AB73D448C5CF CRC64;

Query Match 29.2%; Score 42; DB 8; Length 23;  
 Best Local Similarity 53.3%; Pred. No. 13;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 YERKKRRRRSGSG 15  
 DB 3 FTRRRRRRTNIGSG 17

RESULT 5  
 ID 065331 PRELIMINARY; PRT; 13 AA.  
 AC 065331:  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DE 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 GN BASIC PROTEIN (FRAGMENT).  
 OS Autographa californica nuclear polyhedrosis virus (AcMNPV).  
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
 OC Nucleopolyhedrovirus.  
 OX NCBI\_TaxID=46015;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C6;  
 RX MEDLINE=90218045; PubMed=2109042;  
 RA Hill-Parkins M.S., Possee R.D.;  
 RL "A baculovirus expression vector derived from the basic protein  
 RT promoter of Autographa californica nuclear polyhedrosis virus";  
 RL J. Gen. Virol. 71:971-976(1990).  
 DR EMBL: D00864; BAA00738.1; -;  
 FT NON\_TER  
 SQ SEQUENCE 13 AA; 1626 MW; 6C75B58A749D1414 CRC64;

Query Match 25.7%; Score 37; DB 12; Length 13;  
 Best Local Similarity 70.0%; Pred. No. 40;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 7 RORRRSGSGT 16  
 DB 4 RRRRSSTGT 13

RESULT 6  
090XS7 PRELIMINARY; PRT; 25 AA.  
AC 090XS7;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE RIBOSOMAL PROTEIN I41.  
OS Ictalurus punctatus (Channel catfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Siluriformes; Ictaluridae; Ictalurus.  
OX NCBI\_Taxid=7998;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Patterson A.P., Karst A., Liu Z.J.;  
RT "Translational machinery of channel catfish: II. Complementary DNA and  
RT Expression of the Complete Set of 47 60S Ribosomal Proteins.";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF401598; AAC9170.1; -  
SQ SEQUENCE 25 AA; 3357 MW; 9AD629C69EDC029F CRC64;

Query Match 24.38; Score 35; DB 13; Length 25;  
Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
QY 2 ERKKRRRRRS 12  
:|:|:|:|:|  
Db 14 KRRRRKMRRS 24

RESULT 7  
096XS2 PRELIMINARY; PRT; 25 AA.  
AC 096XS2;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE RIBOSOMAL PROTEIN I41.  
OS Spodoptera frugiperda (Fall armyworm).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
OC Noctuidae; Noctuidae; Amphipyrinae; Spodoptera.  
OX NCBI\_Taxid=7108;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Landais I., Ogilastro M., Mita K., Nohata J., Lopez-Ferber M.,  
RA Duonot-Geruttl M., Fournier P., Devauchelle G.;  
RT "Full-length ribosomal protein sequence from an EST library of  
RT Spodoptera frugiperda cells (Sf9).";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF400204; AAK92176.1; -  
SQ SEQUENCE 25 AA; 3399 MW; 9AD629D9ECC039F CRC64;

Query Match 22.98; Score 33; DB 5; Length 25;  
Best Local Similarity 54.5%; Pred. No. 3.4e+02;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 ERKKRRRRRS 12  
:|:|:|:|:|  
Db 14 KRRRRKMRRS 24  
RESULT 8  
057542 PRELIMINARY; PRT; 27 AA.  
AC 057542;  
DT 01-JUN-1998 (TREMBLrel. 06, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DR 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE HOMEOBOX PROTEIN LPHOX1B (FRAGMENT).  
OS Lampetra planeri (Brook lamprey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;  
OC Petromyzontiformes; Petromyzontidae; Lampetra.  
OX NCBI\_Taxid=7750;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98358009; PubMed=9694633;  
RA Sharnan A.C., Holland P.W.;  
RT "Estimation of Hox gene cluster number in lampreys.";  
RL Int. J. Dev. Biol. 42:617-620(1998).  
DR EMBL: AF044798; AAC03002.1; -  
FT NON-TER 1  
FT NON-TER 27  
SQ SEQUENCE 27 AA; 3098 MW; FE1EE5E971DDCIDA CRC64;

Query Match 22.9%; Score 33; DB 13; Length 27;  
Best Local Similarity 46.7%; Pred. No. 3.8e+02;  
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
QY 3 RKRRRRSSSGTD 17  
|:|:|:|:|:|  
Db 9 RAVRRRRRAAARD 23

RESULT 9  
057543 PRELIMINARY; PRT; 27 AA.  
AC 057543;  
DT 01-JUN-1998 (TREMBLrel. 06, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DE HOMEOBOX PROTEIN LPHOX1C (FRAGMENT).  
OS Lampetra planeri (Brook lamprey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;  
OC Petromyzontiformes; Petromyzontidae; Lampetra.  
OX NCBI\_Taxid=7750;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98358009; PubMed=9694633;  
RA Sharnan A.C., Holland P.W.;  
RT "Estimation of Hox gene cluster number in lampreys.";  
RL Int. J. Dev. Biol. 42:617-620(1998).  
DR EMBL: AF044799; AAC03003.1; -  
FT NON-TER 1  
FT NON-TER 27  
SQ SEQUENCE 27 AA; 3190 MW; B86EE73971DDDC06 CRC64;

Query Match 22.9%; Score 33; DB 13; Length 27;  
Best Local Similarity 46.7%; Pred. No. 3.8e+02;  
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 RKRRRRSSSGTD 17  
|:|:|:|:|:|  
Db 9 RATRRRRRAATERD 23  
RESULT 10  
09TSS7 PRELIMINARY; PRT; 29 AA.  
AC 09TSS7;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE WILM'S TUMOR PROTEIN 1 (FRAGMENT).  
GN WT1.  
OS Felis silvestris catus (Cat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteleia; Carnivora; Placentalia; Felidae; Felis.  
OX NCBI\_Taxid=9685;  
RN [1]

RP	SEQUENCE FROM N.A.	
RX	MEDLINE=21015404; Pubmed=11330975;	
RA	Brouillette J.A., Andrew J.R., Venta P.J.,	
RT	"Estimate of nucleotide diversity in dogs with a pooled	
RT	method."	
RI	Mem. Genome. 11:1079-1086(2000).	
DR	EMBL; AF201740; AAF19825.1; -; JOINED.	
DR	EMBL; AF201735; AAF19825.1; JOINED.	
FT	NON-VER	1
FT	NON-VER	1
Q0	SEQUENCE	29 AA: 3594 MW; FDC8E57A0BAE35DD CRC64;

Query Match	22.28;	Score 32;	DB 6;	Length 29;
Best Local Similarity	66.78;	Pred. NO. 5.8e+02;		
Matches	6;	Conservative	1;	Mismatches 2;
			Indels	0;
			Gaps	0;

```
QY      5 KRRQRRSG 13
        ||||| :|
Db      1 KRNQRHTG 9
```

RESULT	11
085496	
AC 085496:	PRELIMINARY; PRT; 23 AA.
DT 01-NOV-1996 (TEMBLrel_01, Created)	
DT 01-NOV-1996 (TEMBLrel_01, Last sequence update)	
DT 01-DEC-2001 (TEMBLrel_19, Last annotation update)	
DE ENVELOPE PROTEIN (FUSION)	
OS Bovine leukemia virus (BLV).	
OC Viruses; Retrofold viruses; Retroviridae; Deltaretrovirus.	
OX NCBI_Taxid=11901;	
RN [1]	
RP SEQUENCE FROM N.A.	
RA MEDLINE=87198886; Pubmed=3033284;	
RA Rice N.R., Slineck S.L., Dubois G.C., Showalter S.D., Gilden R.V.,	
RA Stephens R.M.: The bovine leukemia virus x region in virus-infected	
RT "expression of the bovine leukemia virus x region in virus-infected	
RT cells."	
RT J. VIROL. 61:1577-1585(1987).	
RL EMBL: M16017; AAA87336.1; -.	
DR NON_TER 23	
FO SEQUENCE 23 AA: 287 MW; F85E7CBED5440B08 CRC64;	

Query Match	21.5%;	Score 31;	DB 15;	Length 23;
Best Local Similarity	75.0%;	Pred. No. 6.4e+02;		
Matches	6;	Conservative	1;	Mismatches 1;
				Indels 0;
				Gaps 0;

```
QY      4 KKRQRRR 11
        1:11111
Db      3 KERRRRR 10
```

RESULT	12			
094781				
ID	094781	PRELIMINARY:	PRT;	23 AA.
AC	094781;			
DT	01-FEB-1997	(TEMBLrel. 02, Created)		
DT	01-FEB-1997	(TEMBLrel. 02, Last sequence update)		
DT	01-JAN-1999	(TEMBLrel. 09, Last annotation update)		
DE	HISTONE H2A (FRAGMENT).			
OS	Trypanosoma cruzi.			
OC	Eukaryota; Eulenzozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma			
OX	NCBI_TaxID=5693;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-Y;			
RA	Tanaka T., Tanaka M.;			
RT	"The parasite genome project: Trypanosoma cruzi.";			
RL	Submitted (Aug-1996) to the EMBL/GenBank/DBD databases.			
EMBL	DB7227; BAA13318.1; -			

```

FT      NON_TIER      1      1
SQ      SEQUENCE      23 AA; 2790 MW; 12E9ED7592E52045 CRC64;

```

Query Match	20.8%	Score 30;	DB 5;	Length 23;
Best Local Similarity	66.7%	Pred. No. 9.1e+02;		
Matches	6;	Conservative	2;	Indels 0;
				Gaps 0;

```
QY      3 RKKRRQRRR 11
        | | | | |
Db     10 RDKRRKRER 18
```

RESULT	13
042364	
ID	042364
AC	042364
DC	01-NOV-1996 (TREMBLER, 01, Created)
DT	01-NOV-1996 (TREMBLER, 01, last sequence update)
DT	01-OCT-2000 (TREMBLER, 15, last annotation update)
DE	ALPHA-AMYLASE SUBFAMILY AMY1 PROTEIN (FRAGMENT).
OS	Triticum aestivum (Wheat).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
CC	Triciteae; Triticum.
OX	NCBI_TaxID=4565;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=9236494; Pubmed=1502164;
RA	Huang N., Stebbins G.L., Rodriguez R.L.;
RT	"Classification and evolution of alpha-amylase genes in plants.";
RL	Proc. Natl. Acad. Sci. U.S.A. 89:7526-7530(1992).
DR	EMBL: S82213; AAD13822.1; -.
DR	HSSP: P04063; IAVA.
FT	NON_TER 1
SO	SEQUENCE 28 AA; 3022 MW; 65EE0A4138FA3A1 CRC64;

Query Match	20.8%;	Score 30;	DB 10;	Length 28;
Best Local Similarity	42.9%;	Pred. No. 1.1e+03;		
Matches	6;	Conservative	1;	Mismatches 7;
			Indels	0;
			Gaps	0;

```
QY      12 SGGSGDTFTLTISL 25
          |||||
Db      14 SGPCTTFDFTTKGI 27
```

RESULT	14			
Q919P3				
ID	Q919P3	PRELIMINARY:	PRT:	22 AA.
AC	Q919P3;			
DT	01-OCT-2000 (Tremblrel. 15, Created)			
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)			
DT	01-OCT-2000 (Tremblrel. 15, Last annotation update)			
DE	CHROMO-HELICASE-DNA BINDING PROTEIN (FRAGMENT).			
OS	Ara miltaris.			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Enchrasauria; Aves; Neognathae; Psittaciformes; Psittacidae;			
OX	NCBI_TaxID=57237;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RA	Leal-Garcia C.H., Bermudez-Humarán L.G., Jaramillo-Rangel G.,			
RA	Rojas-Valdes V., Montes de Oca-Luna R.;			
RT	*Molecular sexing of monomorphic endangered Ara birds.*;			
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF229847; AAF42915.1; -			
KW	Helicase.			
FT	NON_TER	1	1	
FT	NON_TER	22	22	
SQ	SEQUENCE	22 AA;	2703 MW;	CBEJ13B31A56673 CRC64;

Query Match	20.18; Score 29; DB 13; Length 22;
-------------	------------------------------------



Best Local Similarity 60.0%; Pred. No. 1.2e+03;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 ERKRRQRR 11  
11:1:1  
Db 4 ERKRRRGR 13

RESULT 15

ID 019358 PRELIMINARY; PRT; 28 AA.  
AC 019358;  
DT 01-JAN-1998 (TREMblrel. 05, Created)  
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE BOTA PROTEIN (FRAGMENT).  
GN BOTA.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93052564; PubMed=1428011;  
RA Ellis S.A., Breem K.A., Morrison W.I.;  
RT "Transmembrane and cytoplasmic domain sequences demonstrate at least  
two expressed bovine MHC class I loci."  
RL Immunogenetics 37:49-56(1992).  
DR EMBL: S47737; AAB23971.1; -.  
FT NON\_TPR 1  
SQ SEQUENCE 28 AA; 2852 MW; 78798D5C19C16CF4 CRC64;

Query Match 20.1%; Score 29; DB 7; Length 28;  
Best Local Similarity 45.5%; Pred. No. 1.6e+03;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 12 SCGGTDFFTI 22  
1:1:1:1:  
Db 15 SAQGSVDVSLTV 25

Search completed: July 15, 2002, 13:57:48  
Job time: 336 sec





PS Disclosure; Page 18; 157pp; English.

XX  
XX A method has been developed for the diagnosis of human Rheumatoid  
CC arthritis (RA) using two-dimensional electrophoresis to generate a  
CC two-dimensional array of features. The method can be used for screening,  
CC diagnosis and prognosis of RA in a subject or for monitoring the effect  
CC of an anti-RA drug or therapy administered to a subject. The method  
CC comprises: (a) analysing a sample of serum or plasma and optionally  
CC synovial fluid by two-dimensional electrophoresis, to generate a two-  
CC dimensional array of features; (b) identifying at least one chosen  
CC feature whose relative abundance correlates with the presence or absence  
CC of RA; and (c) comparing the abundance of each chosen feature in the  
CC sample with the abundance of that chosen feature in serum or plasma from  
CC one or more persons without RA, where the relative abundance of the  
CC chosen feature or features in the sample indicates the presence or  
CC absence of RA in the subject. The method can also be used in clinical  
CC studies for testing drugs for therapy of RA, for purification of RA-  
CC diagnostic protein isoforms (RPIs) and for production of antibodies to  
CC RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify  
CC compounds that promote or inhibit their activity, which are then used as  
CC RA drugs. Nucleic acid encoding RADFs can be used in gene therapy  
CC protocols. AA141844 to AA142100 represent RPI peptides, AA142101 to  
CC AA142103 represent expression reference protein isoform peptides and  
CC AA25066 to AA25068 represent degenerate probes for RPIs, which are all  
CC used in the exemplification of the present invention.

XX  
XX Sequence 16 AA:

QY 12 SSGSGTDFTLTIS 23  
Db 4 ssgstdfiltis 15

Query Match 41.0%; Score 59; DB 20; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.089;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2  
AAU25389  
ID AAU25389 standard; Peptide: 16 AA.

XX AAU25389;  
AC  
XX 18-DEC-2001 (first entry)  
DT  
XX  
XX Schizophrenia-Associated Protein Isoform (SPI) peptide #618.  
DE  
XX  
XX Schizophrenia-associated protein isoform; SPI; SPI-206; SPI-238; SPI-240;  
KW neuroleptic; gene therapy; cerebrospinal fluid; serum; plasma.  
OS  
XX Homo sapiens.  
OS  
XX  
XX WO200162785-A2.  
PN  
XX  
XX 30-AUG-2001.  
PD  
XX  
XX 23-FEB-2001; 2001WO-GB00792.  
PE  
XX  
XX 24-FEB-2000; 2000GB-0004415.  
PR 28-NOV-2000; 2000US-0750395.  
XX  
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
PA  
XX Herath HMAc, Parekh RB, Rohlf C, Terrett JA, Tyson KL;  
XX Herath HMAc, Parekh RB, Rohlf C, Terrett JA, Tyson KL;  
XX WPI; 2001-570624/64.  
DR  
XX  
XX New schizophrenia associated protein isoforms and encoding nucleic acid  
PT molecules, useful for treatment, diagnosis and prognosis of  
PT schizophrenia and screening for potential drugs for treatment and new  
PT drug targets -  
XX

PS Disclosure; Page 41; 146pp; English.

XX  
XX The sequence represents a schizophrenia-associated protein isoform (SPI).  
CC These protein isoforms, e.g. SPI-206, SPI-238 and SPI-240 are detectable  
CC in cerebrospinal fluid, serum or plasma and are useful markers of  
CC schizophrenia. The sequences can be used for treatment and diagnosis of  
CC schizophrenia, screening, prognosis, monitoring the results of therapy,  
CC identifying patients most likely to respond to a particular therapy and  
CC identification of new targets for drug treatment. SPI DNA is useful as a  
CC nucleic acid probe to detect the presence of nucleic acids or SPIs.

XX  
XX Sequence 16 AA:

QY 12 SSGSGTDFTLTIS 23  
Db 4 ssgstdfiltis 15

Query Match 41.0%; Score 59; DB 22; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.089;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3  
AAU15733  
ID AAU15733 standard; Peptide: 16 AA.

XX AAU15733;  
AC  
XX 24-OCT-2001 (first entry)  
DT  
XX  
XX Schizophrenia-associated isoform peptide #618.  
DE  
XX  
XX Schizophrenia; neuroleptic; diagnostic; neuropsychiatric disorder;  
KW neurological disorder; neuropathy.  
OS  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200163293-A2.  
PN  
XX  
XX 30-AUG-2001.  
PD  
XX  
XX 23-FEB-2001; 2001WO-GB00783.  
PE  
XX  
XX 24-FEB-2000; 2000GB-0004415.  
PR 28-NOV-2000; 2000US-0750395.  
XX  
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
PA  
XX Herath HMAc, Parekh RB, Rohlf C;  
XX Herath HMAc, Parekh RB, Rohlf C;  
XX WPI; 2001-502868/55.  
DR  
XX  
XX Diagnosing and monitoring Schizophrenia by detecting the presence of  
PT Schizophrenia Associated features and Schizophrenia Associated Protein  
PT isoforms in samples of cerebrospinal fluid -  
XX  
XX  
XX Claim 6; Page 41; 160pp; English.

XX  
XX The invention relates to methods and compositions for screening,  
CC diagnosis and prognosis of Schizophrenia. The method involves detecting  
CC the presence of Schizophrenia (SCH) Associated features (SfEs) and SCH  
CC Associated Protein Isoforms (SPIs) in samples, e.g. by electrophoresis,  
CC immunoassay or hybridisation assay, for diagnosing and monitoring SCH,  
CC studying the effectiveness of treatments and for identifying potential  
CC therapeutic agents. The method is used for (1) screening or diagnosis of  
CC SCH and the relative abundance of at least 1 chosen feature correlates  
CC with the presence or absence of SCH; and (2) monitoring the effect of  
CC therapy administered to a subject with SCH and the relative abundance of  
CC at least 1 chosen feature which correlates with the severity of SCH.  
CC The expression and activity of the SfEs, SPIs and related molecules  
CC (e.g. secondary messengers) are studied to diagnose SCH, monitor the  
CC progress of the disorder and the effectiveness of treatment and as

CC targets to identify and produce potential therapeutic agents for the  
CC treatment of SCH. The paucity of detectable neuralgic defects  
CC distinguishes neuropsychiatric disorders such as SCH from neurological  
CC disorders, where manifestations of anatomical and biochemical changes  
CC have been identified in many cases. Consequently the identification and  
CC characterisation of cellular and/or molecular causative defects and  
CC neuropathies are necessary for improved treatment of neuropsychiatric  
CC disorders. AAU15114-AAU15762 represent the amino acid sequences of  
CC schizophrenia-associated isoforms used in the method of the invention.  
XX  
XX Sequence 16 AA:

Query Match 41.0%; Score 59; DB 22; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.089;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 SGGSTDPFTLTIS 23  
Db 4 sgsqtdtltis 15

## RESULT 4

AA86601 ID AAR86601 standard; peptide; 18 AA.

XX AC AAR86601:

XX DT 28-JUN-1996 (first entry)

XX DE HIV TAT cellular uptake region immunogen #1.

XX KW Immunogen; cellular uptake region; transactivating protein; TAT; HIV;  
XX KX viraemia; antibody.

XX OS Human immunodeficiency virus type 1.

XX PN MO9531999-A1.

XX PD 30-NOV-1995.

XX PF 16-MAY-1995; 95WO-US06077.

XX PR 23-MAY-1994; 94US-0247991.

XX PA (TMU-) IMMUNOCITOLOGY RES INST INC.

XX PI Culler MD, Goldstein G, Shenbagamurthi P;

XX DR WPI; 1996-049298/05.

XX PT Immunogen containing cellular uptake region of viral TAT protein -  
XX PT induces high antibody titre against TAT, partic. for preventing or  
XX PT treating HIV infection

XX PS Claim 6; Page 45; 54pp; English.

XX CC The sequences given in AAR86601-19 are immunogens derived from the  
XX CC cellular uptake region of the transactivating protein, TAT of HIV.  
XX CC These peptides are based on the region comprising amino acids 46-63  
XX CC (see AAR86625) of the HIV TAT protein. These peptides are used to  
XX CC protect against infection by HIV. They may also be used to reduce  
XX CC viraemia in already infected patients. These immunogens may be used  
XX CC to identify, or generate, antibodies specific for TAT.

XX SQ Sequence 18 AA:

Query Match 38.9%; Score 56; DB 17; Length 18;

Best Local Similarity 68.8%; Pred. No. 0.27;  
Matches 11; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 YERKKRRRRSGSGT 16

Db 1 |||||: 1;  
2 ygrkkrrrrrpaggs 17

## RESULT 5

AA49577 ID AAR49577 standard; peptide; 15 AA.

XX AC AAR49577;

XX DT 23-SEP-1994 (first entry)

XX DE Residues tat47-58GCC HIV tat protein.

XX KW HIV; tat; transport; moiety; conjugate; cargo molecule;

XX KM cytoplasmic delivery; nuclear delivery; cysteine-rich region;

XX KW transactivation; disulphide aggregation.

XX OS Human immunodeficiency virus type 1.

XX PN WO9404686-A.

XX PD 03-MAR-1994.

XX PF 19-AUG-1993; 93WO-US07833.

XX PR 21-AUG-1992; 92US-0934375.

XX PA (BIO) BIOGEN INC.

XX PI Barscum JG, Fawell SE, Pepsinsky RB;

XX DR WPI; 1994-083202/10.

XX PT Transport of cargo moieties into cells - using genetic fusions of  
XX PT chemical conjugates comprising a portion of HIV tat protein as  
XX PT transport peptide

XX PS Disclosure; Page 87; 153pp; English.

XX CC The sequences given in AAR49574-77 are fragments of the HIV tat protein  
XX CC which were used as transport moieties in the conjugate of the  
XX CC invention. These conjugates are used as cargo molecules for the  
XX CC efficient cytoplasmic and nuclear delivery of biologically active  
XX CC non-tat proteins, nucleic acids and other molecules which are not  
XX CC inherently capable of entering a target cell. The tat basic region  
XX CC amino acid sequence carries cargo molecules by covalent attachment.  
XX CC The reduced size of the transport peptides minimises interference  
XX CC with the biological activity of the cargo molecule. By virtue of  
XX CC the absence of the cysteine-rich region (residues 22-36) of the tat  
XX CC protein, problems of spurious transactivation and disulphide  
XX CC aggregation are solved. The reduced transport peptide size also  
XX CC enhances uptake efficiency.

XX SQ Sequence 15 AA:

Query Match 38.2%; Score 55; DB 15; Length 15;

Best Local Similarity 84.6%; Pred. No. 0.31;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YERKKRRRRRSG 13

Db 1 ygrkkrrrrrrpp 13

## RESULT 6

AAW31207 ID AAW31207 standard; peptide; 15 AA.

XX AC AAW31207;

XX DT 20-MAR-1998 (first entry)

```
XX DE HIV-1 based transport polypeptide Tat47-58GCC.
XX XX
XX KW Human immunodeficiency virus; HIV Type 1; Tat protein;
XX KW cargo molecules; intracellular delivery; fusion protein;
XX KW therapeutic; prophylactic; diagnostic; transport polypeptide.
XX OS
XX OS Synthetic.
XX OS Human immunodeficiency virus type 1.
XX
XX FH Key Location/Qualifiers
XX FT Region 3..11
XX FT /label= basic_region
XX FT Misc-difference 13..15
XX FT /label= spacer_residues
XX
XX PN US5674980-A.
XX
XX PD 07-OCT-1997.
XX
XX PP 21-DEC-1989; 89US-0454450.
XX
XX PR 28-APR-1994; 94US-0235403.
XX PR 21-DEC-1989; 89US-0454450.
XX PR 02-JAN-1991; 91US-0636662.
XX PR 21-AUG-1992; 92US-0934375.
XX PR 19-AUG-1993; 93MO-US07833.
XX PR 24-NOV-1993; 93US-0158015.
XX PR 25-MAY-1995; 95US-0450098.
XX
XX PA (BAR/) BARSOUM J G.
XX PA (FAME/) FAWELL S E.
XX PA (FRAN/) FRANKEL A.
XX PA (PABO/) PABO C.
XX PA (PEPI/) PEPINSKY R B.
XX
XX PI Barsoum JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;
XX DR WPI: 1997-502388/46.
XX
XX PT Fusion proteins containing truncated HIV tat sequences - useful for
XX PT intracellular delivery of viral repressor proteins
XX
XX Example 1: Column 55; 77pp; English.
XX
XX PS This sequence represents the transport polypeptide Tat47-58GCC which can
XX PS be used in a novel method for the delivery of biologically active cargo
XX PS molecules into the cytoplasm and nuclei of eukaryotic cells. The tat
XX PS protein from immunodeficiency virus (e.g. HIV-1, HIV-2, STV) is readily
XX PS taken up into cells when present extracellularly. This polypeptide
XX PS has been constructed from the HIV-1 tat protein (see AAW31202) by
XX PS removing the cysteine-rich region and the carboxy-terminal domain while
XX PS retaining the basic region. Cargo molecules could be covalently linked
XX PS to the protein to form a fusion protein without the existing problems of
XX PS spurious trans-activation and disulphide aggregation. The reduced size
XX PS of transport polypeptides also minimises interference with the
XX PS biological activity of the cargo molecule. This is applicable for
XX PS therapeutic, prophylactic or diagnostic intracellular delivery of small
XX PS molecules and macromolecules e.g. proteins, nucleic acids and
XX PS polysaccharides.
XX
XX SQ Sequence 15 AA;
```

```
Query Match 38.2%; Score 55; DB 18; Length 15;
Best Local Similarity 84.6%; Pred. No. 0.31;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 1 YEKKKRRRRRSG 13
DB 1 Ygrkkrrrrrrpg 13
```

```
RESULT 7
AAW32916
XX ID AAW32916 standard; peptide; 15 AA.
XX
XX AC AAW32916;
XX
XX DT 19-JAN-1998 (first entry)
XX
XX DE HIV-1 tat protein residues 47-58 plus Gly-Gly-Cys linker.
XX
XX KW Chemical conjugate; transport polypeptide; tat protein;
XX KW nucleic acid; delivery; cytoplasm; cell nucleus; therapy;
XX KW prophylaxis; diagnosis; spurious transactivation; HIV-1;
XX KW disulphide aggregation.
XX
XX OS Human immunodeficiency virus type 1.
XX
XX PN US5670617-A.
XX
XX PD 23-SEP-1997.
XX
XX PP 21-DEC-1989; 89US-0454450.
XX
XX PR 28-APR-1994; 94US-0235403.
XX PR 21-DEC-1989; 89US-0454450.
XX PR 02-JAN-1991; 91US-0636662.
XX PR 21-AUG-1992; 92US-0934375.
XX PR 19-AUG-1993; 93MO-US07833.
XX PR 24-NOV-1993; 93US-0158015.
XX PR 25-MAY-1995; 95US-0450246.
XX
XX PA (BAR/) BARSOUM J G.
XX PA (FAME/) FAWELL S E.
XX PA (FRAN/) FRANKEL A.
XX PA (PABO/) PABO C.
XX PA (PEPI/) PEPINSKY R B.
XX
XX PI Barsoum JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;
XX DR WPI: 1997-479523/44.
XX
XX PT Conjugate for intracellular delivery - comprising transport moiety
XX PT having amino acids 49-57 of human immunodeficiency virus tat protein
XX PT and nucleic acid cargo moiety
XX
XX PS Disclosure: Columns 57-58; 77pp; English.
XX
XX PS A novel covalently linked chemical conjugate, comprises a transport
XX PS polypeptide moiety consisting of at least residues 49-57 of human
XX PS immunodeficiency virus (HIV) tat protein, but not residues 22-36
XX PS and 73-86, e.g. the present peptide, and a cargo moiety comprising
XX PS a single or double stranded nucleic acid. The conjugate can be used
XX PS to deliver cargo molecules into the cytoplasm and nuclei of cells
XX PS for therapeutic, prophylactic and diagnostic applications. The HIV
XX PS tat protein is readily taken up into cells and the cell nucleus.
XX PS The reduced size of the transport polypeptides minimises
XX PS interference with the biological activity of the cargo molecule.
XX PS In addition, by virtue of the absence of the cysteine rich region
XX PS of the HIV tat protein, the transport polypeptides solve the
XX PS potential problems of spurious transactivation and disulphide
XX PS aggregation.
XX
XX SQ Sequence 15 AA;
```

```
Query Match 38.2%; Score 55; DB 18; Length 15;
Best Local Similarity 84.6%; Pred. No. 0.31;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 1 YEKKKRRRRRSG 13
DB 1 Ygrkkrrrrrrpg 13
```

```

RESULT 8
AAW26448
ID AAW26448 standard; Protein: 15 AA.
XX
XX AAW26448:
AC
XX
XX 16-DEC-1997 (first entry)
DT
XX
XX HIV tat protein fragment (aa47-58GCC).
DE
XX
XX HIV; tat protein; transport protein; cargo delivery.
KM
XX
XX Chimeric human immunodeficiency virus type 1;
OS
XX Chimeric synthetic.
OS
XX US5652122-A.
XX
XX 29-JUL-1997.
PD
XX
XX 21-DEC-1989; 89US-0454450.
PF
XX
XX 28-APR-1994; 94US-0235403.
PR
XX 21-DEC-1989; 89US-0454450.
PR 02-JAN-1991; 91US-0636662.
PR 21-AUG-1992; 92US-0934375.
PR 19-AUG-1993; 93WO-US07833.
PR 24-NOV-1993; 93US-0158015.
PR 25-MAY-1995; 95US-0450257.
XX
XX (BARS/) BARSON J G.
PA (FRAME/) FAWELL S E.
PA (FRAN/) FRANKEL A.
PA (PABO/) PABO C.
PA (PEPI/) PEPIISKY R B.
PI
XX Barsom JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;
XX
XX WPI: 1997-392943/36.
DR
XX
XX New DNA constructs for transporting molecules to cells - encode a
PT fusion protein comprising a modified HIV tat protein and a
PT carboxy-terminal cargo moiety
XX
XX Disclosure: Column 57-58; 76pp; English.
XX
XX This polypeptide comprises amino acids 47-58 of the HIV tat
CC protein (see also AAW26443), plus a Gly-Gly-Cys C-terminal extension.
CC Novel transport proteins comprise modified HIV tat covalently
CC attached to a cargo molecule (see AAW26436-42). The modified tat
CC protein (see AAW26444-49) is characterised by the presence of the tat
CC basic region (amino acids 47-59), and the absence of the Cys-rich
CC region (amino acids 22-36) and the exon 2-encoded C-terminal region
CC (amino acids 73-86). These modifications solve the potential
CC problems of spurious trans-activation and disulphide aggregation,
CC while the reduced size of the transport protein minimises
CC interference with the biological activity of the cargo molecule.
CC DNA molecules that encode the modified tat fusion proteins are
CC claimed and can be used to deliver polypeptides or nucleic acids to
CC the cytoplasm or nuclei of cells in vivo or in vitro.
XX
XX
XX Sequence 15 AA:
SO

```

Query Match 38.2%; Score 55; DB 18; Length 15;  
 Best Local Similarity 84.6%; Pred. No. 0.31;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

OY 1 YERKKRRORRRSG 13
   | | | | | | | | | |
Db 1 ygrkkrrgrrrpg 13

```

```

RESULT 9
AAW76153
ID AAW76153 standard; protein: 15 AA.
XX
XX AAW76153:
AC
XX
XX 24-NOV-1998 (first entry)
DT
XX
XX HIV Type I TAT protein fragment tat47-58GCC.
DE
XX
XX TAT protein; cargo molecule; therapy; diagnosis; transport protein;
KM fusion protein; human papillomavirus E2 repressor; target cell.
XX
XX Human immunodeficiency virus type 1.
OS
XX US5804604-A.
XX
XX 08-SEP-1998.
PD
XX
XX 25-MAY-1995; 95US-0450236.
PF
XX
XX 28-APR-1994; 94US-0235403.
PR 21-DEC-1989; 89US-0454450.
PR 02-JAN-1991; 91US-0636662.
PR 19-AUG-1993; 93WO-US07833.
PR 24-NOV-1993; 93US-0158015.
PR 25-MAY-1995; 95US-0450236.
XX
XX (BIOI ) BIOGEN INC.
PA
PI
XX Barsom JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;
XX
XX WPI: 1998-505702/43.
DR
XX
XX HIV tat-derived transport fusion proteins - used to deliver
PT biological active molecules e.g. peptide(s) or nucleic acids,
PT specifically into cytoplasm or nuclei of cells
XX
XX Disclosure: Column 57-58; 83pp; English.
XX
XX This sequence represents a human immunodeficiency virus (HIV) Type I
CC TAT protein fragment, tat47-58GCC which is used in a method for the
CC delivery of biologically active cargo molecules into the cytoplasm and
CC nuclei of cells, for therapeutic, prophylactic or diagnostic purposes.
CC This is accomplished by the presence of a small, basic section of tat
CC transport protein of HIV. This is used as it is this protein which is
CC observed to cause human cells in culture to take up HIV. The method
CC involves the use of a cargo moiety in combination with a transport
CC moiety usually in the form of a fusion protein. The cargo moiety is a
CC human papillomavirus E2 repressor that retains its biological activity
CC after delivery into a target cell and where the transport moiety is
CC one of following HIV tat protein fragments (a) aa 47-58, (b) aa 47-72,
CC (c) 38-72, (d) aa 38-58, (e) aa 37-58, (f) aa 1-21 and 38-72.
CC (g) aa 47-62 or aa 38-62. The proteins allow delivery of specific
CC peptides into cells at high concentrations due to use of existing
CC transporters. Previous methods of delivery include bombardment and
CC transforming, which only allow a fraction of the cell population to be
CC infected and can additionally damage cells as they cause physical opening
CC of the cell walls/membranes to allow entry.
XX
XX
XX Sequence 15 AA:
SO

```

Query Match 38.2%; Score 55; DB 19; Length 15;  
 Best Local Similarity 84.6%; Pred. No. 0.31;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

OY 1 YERKKRRORRRSG 13
   | | | | | | | | | |
Db 1 ygrkkrrgrrrpg 13

```

RESULT 10

AAW19818  
ID AAW19818 standard; Peptide: 16 AA.  
XX  
AC AAW19818;  
XX  
DT 26-JAN-1998 (first entry)  
XX  
DE Universal transfer vector sequence of chimeric coat protein.  
XX  
KW Adenovirus; vector; coat protein; gene therapy; gene transfer;  
KW human; cancer; autoimmune disease; heart disease; infection;  
KW Universal transfer vector; PBSS 75-100 pgs(lat).  
XX  
OS Synthetic.  
XX  
PN W09720051-A2.  
XX  
PD 05-JUN-1997.  
XX  
PF 27-NOV-1996; 96NO-US19150.  
XX  
PR 21-AUG-1996; 96US-0701124.  
PR 28-NOV-1995; 95US-0563368.  
PR 21-AUG-1996; 96US-0700846.  
XX  
PA (GENV-) GENVEC INC.  
XX  
PI Brough DE, Kovesdi I, Wickham TJ.  
XX  
DR WPI: 1997-310606/28.  
DR N-PSDB: AAT70272.  
XX  
PT Adenoviral vectors containing chimeric coat protein - bind and enter  
PT cells more efficiently, useful for gene therapy of e.g. cancer,  
PT autoimmune diseases, etc.  
XX  
PS Claim 7; Page 59-60; 121pp; English.  
XX  
CC This peptide can be used as a universal transfer vector (UTV)  
CC and/or spacer sequence in novel chimeric adenovirus coat proteins  
CC (CPS). The peptide is encoded by a gene fragment (see AAT70272) of  
CC transfer plasmid PBSS 75-100 pgs(lat). Claimed UTVs/spacers are  
CC given in AAW19810-11, AAW19813-25, AAW19827, AAW19829, AAW19831-32 and  
CC AAW19834-43). Claimed CPS, such as fibre proteins, differ from wild-  
CC type CPS by the introduction of the UTV and/or spacer at or near  
CC the C-terminus. This imparts on the chimeric CP the ability to  
CC bind to and enter cells by means of a novel cell surface binding  
CC site. Adenoviral vectors comprising the chimeric CP are able to  
CC enter cells more efficiently than vectors comprising wild-type CP,  
CC especially at lower m.o.i. They are especially useful for gene  
CC therapy of e.g. cancers, genetic disorders, pathogenic infections,  
CC heart disease or autoimmune diseases.  
XX  
SQ Sequence 16 AA:

Query Match 38.2%; Score 55; DB 18; Length 16;  
Best Local Similarity 78.6%; Pred. NO. 0.33;  
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YERKKRRRRRSGS 14  
| | | | | | | | | | | | | | | |  
Db 3 ygrkkrrrrrrrras 16

RESULT 11  
AAE12486  
ID AAE12486 standard; peptide: 22 AA.  
XX  
AC AAE12486;  
XX  
DT 03-JAN-2002 (first entry)  
XX

DE Membrane transiting antiviral peptide bPAT-9.  
XX  
KW Membrane transiting peptide; virucide; antiviral; Herpes Simplex Virus;  
KW HSV; HIV; Human Immunodeficiency Virus; CMV; cytomegalovirus.  
XX  
OS Undefined.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 1 /note="Biotin-aminohexanoyl Tyr"  
XX  
PN W0200157072-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 06-FEB-2001; 2001MO-US03813.  
XX  
PR 07-FEB-2000; 2000US-180823P.  
PR 22-FEB-2000; 2000US-184057P.  
XX  
PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX  
PI Brandt C, Bulmann H;  
XX  
DR WPI: 2001-638840/73.  
XX  
PT Peptides comprising membrane transiting peptides useful for treating or  
PT preventing a virus infection, e.g., human immunodeficiency virus,  
PT herpes simplex virus and cytomegalovirus -  
XX  
PS Claim 7; Page 15; 43pp; English.  
XX  
CC The invention relates to peptides comprising membrane transiting  
CC peptides with antiviral properties. The peptides are useful for  
CC treating or preventing a virus infection in a warm blooded animal,  
CC e.g., enveloped viruses such as human immunodeficiency virus (HIV),  
CC herpes simplex virus (HSV), cytomegalovirus (CMV) and non-enveloped  
CC virus. Preferably, the peptides are useful for treating or preventing  
CC infections from one or more HSVs. The antiviral peptides are used for  
CC treating viral infections of the skin or part of the oral or genital  
CC cavity. The present sequence is membrane transiting antiviral peptide.  
XX  
SQ Sequence 22 AA:

Query Match 38.2%; Score 55; DB 22; Length 22;  
Best Local Similarity 84.6%; Pred. NO. 0.46;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YERKKRRRRRSG 13  
| | | | | | | | | | | | | | | |  
Db 1 ygrkkrrrrrrpg 13

RESULT 12  
AAE12487  
ID AAE12487 standard; peptide: 22 AA.  
XX  
AC AAE12487;  
XX  
DT 03-JAN-2002 (first entry)  
XX  
DE Membrane transiting antiviral peptide bPAT-9x.  
XX  
KW Membrane transiting peptide; virucide; antiviral; Herpes Simplex Virus;  
KW HSV; HIV; Human Immunodeficiency Virus; CMV; cytomegalovirus.  
XX  
OS Undefined.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 1 /note="Biotin-aminohexanoyl Tyr"  
XX



PN WO200157072-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 06-FEB-2001; 2001WO-US03813.  
 XX  
 PR 07-FEB-2000; 2000US-180823P.  
 PR 22-FEB-2000; 2000US-184057P.  
 XX  
 PA (MISC ) WISCONSIN ALUMNI RES FOUND.  
 XX  
 PI Brandt C, Bultmann H;  
 DR WPI; 2001-638840/73.  
 XX  
 PT Peptides comprising membrane transiting peptides useful for treating or  
 PT preventing a virus infection, e.g., human immunodeficiency virus,  
 XX herpes simplex virus and cytomegalovirus -  
 PS Claim 7; Page 15; 43pp; English.  
 XX  
 CC The invention relates to peptides comprising membrane transiting  
 CC peptides with antiviral properties. The peptides are useful for  
 CC treating or preventing a virus infection in a warm blooded animal,  
 CC e.g., enveloped viruses such as human immunodeficiency virus (HIV),  
 CC herpes simplex virus (HSV), cytomegalovirus (CMV) and non-enveloped  
 CC virus. Preferably, the peptides are useful for treating or preventing  
 CC infections from one or more HSVs. The antiviral peptides are used for  
 CC treating viral infections of the skin or part of the oral or genital  
 CC cavity. The present sequence is membrane transiting antiviral peptide.  
 CC  
 SQ Sequence 22 AA:  
 Query Match 38.2%; Score 55; DB 22; Length 22;  
 Best Local Similarity 84.6%; Pred. No. 0.46;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 YERKKRRRRRSG 13  
 DB 1 YRKKRRRRRRSG 13  
 RESULT 13  
 AAR49575  
 ID AAR49575 standard; peptide; 24 AA.  
 XX  
 AC AAR49575;  
 XX  
 DT 23-SEP-1994 (first entry)  
 XX  
 DE Residues 37-58GGC HIV tat protein.  
 XX  
 KW HIV; tat; transport; motety; conjugate; cargo molecule;  
 KW cytoplasmic delivery; nuclear delivery; cysteine-rich region;  
 KW transactivation; disulphide aggregation.  
 XX  
 OS Human immunodeficiency virus type 1.  
 XX  
 PN WO9404686-A.  
 PN  
 PD 03-MAR-1994.  
 XX  
 PF 19-AUG-1993; 93WO-US07833.  
 PR  
 PR 21-AUG-1992; 92US-0934375.  
 XX  
 PA (BIOJ ) BIOGEN INC.  
 XX  
 PI Barsom JG, Fawell SE, Pepinsky RB;  
 XX  
 DR WPI; 1994-083202/10.  
 XX

PT Transport of cargo moieties into cells - using genetic fusions of  
 PT chemical conjugates comprising a portion of HIV tat protein as  
 PT transport peptide  
 XX  
 PS Disclosure; Page 86; 153pp; English.  
 XX  
 CC The sequences given in AAR49574-77 are fragments of the HIV tat protein  
 CC which were used as transport moieties in the conjugate of the  
 CC invention. These conjugates are used as cargo molecules for the  
 CC efficient cytoplasmic and nuclear delivery of biologically active  
 CC non-tat proteins, nucleic acids and other molecules which are not  
 CC inherently capable of entering a target cell. The tat basic region  
 CC amino acid sequence carries cargo molecules by covalent attachment.  
 CC The reduced size of the transport peptides minimises interference  
 CC with the biological activity of the cargo molecule. By virtue of  
 CC the absence of the cysteine-rich region (residues 22-36) of the tat  
 CC protein, problems of spurious transactivation and disulphide  
 CC aggregation are solved. The reduced transport peptide size also  
 CC enhances uptake efficiency.  
 CC  
 SQ Sequence 24 AA:  
 Query Match 38.2%; Score 55; DB 15; Length 24;  
 Best Local Similarity 84.6%; Pred. No. 0.5;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 YERKKRRRRRSG 13  
 DB 10 YRKKRRRRRRSG 22  
 RESULT 14  
 AAW31205  
 ID AAW31205 standard; peptide; 24 AA.  
 XX  
 AC AAW31205;  
 XX  
 DT 20-MAR-1998 (first entry)  
 XX  
 DE HIV-1 based transport polypeptide Tat38-58GGC.  
 XX  
 KW Human immunodeficiency virus; HIV Type 1; Tat protein;  
 KW cargo molecules; intracellular delivery; fusion protein;  
 KW therapeutic; prophylactic; diagnostic; transport polypeptide.  
 XX  
 OS Synthetic.  
 OS Human immunodeficiency virus type 1.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 12..20  
 FT /Label= basic\_domain  
 FT Misc-difference 22..24  
 FT /note= "spacer residues"  
 XX  
 PN US5674980-A.  
 XX  
 PD 07-OCT-1997.  
 XX  
 PF 21-DEC-1989; 89US-0454450.  
 XX  
 PR 28-APR-1994; 94US-0235403.  
 PR 21-DEC-1989; 89US-0454450.  
 PR 02-JAN-1991; 91US-0636662.  
 PR 21-AUG-1992; 92US-0934375.  
 PR 19-AUG-1993; 93WO-US07833.  
 PR 24-NOV-1993; 93US-0158015.  
 PR 25-MAY-1995; 95US-0450098.  
 XX  
 PA (BARS/) BARSCUM J G.  
 PA (FAWE/) FAWELL S E.  
 PA (FRAN/) FRANKEL A.  
 PA (PABO/) PABO.

PA (PEPI/) PEPINSKY R B.  
 XX  
 PI Barsom JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;  
 XX WPI: 1997-502388/46.  
 DR  
 XX  
 PT Fusion proteins containing truncated HIV tat sequences - useful for  
 PT intracellular delivery of viral repressor proteins  
 XX  
 PS Example 1; Column 55; 77pp; English.  
 CC This sequence represents the transport polypeptide Tat38-586GC which can  
 CC be used in a novel method for the delivery of biologically active cargo  
 CC molecules into the cytoplasm and nuclei of eukaryotic cells. The tat  
 CC protein from immunodeficiency virus (e.g. HIV-1, HIV-2, SIV) is readily  
 CC taken up into cells when present extracellularly. This polypeptide  
 CC has been constructed from the HIV-1 tat protein (see AAW3102) by  
 CC removing the cysteine-rich region and the carboxy-terminal domain while  
 CC retaining the basic region. Cargo molecules could be covalently linked  
 CC to the protein to form a fusion protein without the existing problems of  
 CC spurious trans-activation and disulphide aggregation. The reduced size  
 CC of transport polypeptides also minimises interference with the  
 CC biological activity of the cargo molecule. This is applicable for  
 CC therapeutic, prophylactic or diagnostic intracellular delivery of small  
 CC molecules and macromolecules e.g. proteins, nucleic acids and  
 CC polysaccharides.  
 CC  
 XX  
 SQ Sequence 24 AA;

Query Match 38.2%; Score 55; DB 18; Length 24;  
 Best Local Similarity 84.6%; Pred. No. 0.5;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 YEKRRRRRRRG 13  
 | | | | | | | | | |  
 DB 10 YGRKKRRQRRTPG 22

RESULT 15  
 AAW32914  
 ID AAW32914 standard; peptide; 24 AA.  
 XX  
 AC AAW32914;  
 XX  
 DT 19-JAN-1998 (first entry)  
 XX  
 DE HIV-1 tat protein residues 38-58 plus Gly-Gly linker.  
 XX  
 KW Chemical conjugate; transport polypeptide; tat protein;  
 KW nucleic acid; delivery; cytoplasm; cell nucleus; therapy;  
 KW prophylaxis; diagnosis; spurious transactivation; HIV-1;  
 KW disulphide aggregation.  
 XX  
 OS Human immunodeficiency virus type 1.  
 XX  
 PN US5670617-A.  
 XX  
 PD 23-SEP-1997.  
 XX  
 PF 21-DEC-1989; 89US-0454450.  
 XX  
 PR 28-APR-1994; 94US-0235403.  
 PR 21-DEC-1989; 89US-0454450.  
 PR 02-JAN-1991; 91US-0636662.  
 PR 21-AUG-1992; 92US-0934375.  
 PR 19-AUG-1993; 93WO-US07833.  
 PR 24-NOV-1993; 93US-0158015.  
 PR 25-MAY-1995; 95US-0450246.  
 XX  
 PA (BAR/) BARSOM J G.  
 PA (FAWE/) FAWELL S E.  
 PA (FRAN/) FRANKEL A.

PA (PABO/) PABO C.  
 PA (PEPI/) PEPINSKY R B.  
 XX  
 PI Barsom JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;  
 XX WPI: 1997-479523/44.  
 DR  
 XX  
 PT Conjugate for intracellular delivery - comprising transport moiety  
 PT having amino acids 49-57 of human immunodeficiency virus tat protein  
 PT and nucleic acid cargo moiety  
 XX  
 PS Disclosure; Columns 55-56; 77pp; English.  
 CC A novel covalently linked chemical conjugate, comprises a transport  
 CC polypeptide moiety consisting of at least residues 49-57 of human  
 CC immunodeficiency virus (HIV) tat protein, but not residues 22-36  
 CC and 73-86, e.g. the present peptide, and a cargo moiety comprising  
 CC a single or double stranded nucleic acid. The conjugate can be used  
 CC to deliver cargo moieties into the cytoplasm and nuclei of cells  
 CC for therapeutic, prophylactic and diagnostic applications. The HIV  
 CC tat protein is readily taken up into cells and the cell nucleus.  
 CC The reduced size of the transport polypeptides minimises  
 CC interference with the biological activity of the cargo molecule.  
 CC In addition, by virtue of the absence of the cysteine rich region  
 CC of the HIV tat protein, the transport polypeptides solve the  
 CC potential problems of spurious transactivation and disulphide  
 CC aggregation.  
 CC  
 XX  
 SQ Sequence 24 AA;

Query Match 38.2%; Score 55; DB 18; Length 24;  
 Best Local Similarity 84.6%; Pred. No. 0.5;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 YEKRRRRRRRG 13  
 | | | | | | | | | |  
 DB 10 YGRKKRRQRRTPG 22

Search completed: July 15, 2002, 13:51:41  
 Job time: 230 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:49:36 ; Search time 21.58 Seconds

(without alignments)  
32.824 Million cell updates/sec

Title: US-09-712-819a-9

Perfect score: 144  
Sequence: 1 YEKRRRRRRSSGSDFTLTITSSLOAED 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 142465

Minimum DB seq length: 0  
Maximum DB seq length: 29

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCNUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	79	54.9	25	5	PCT-US91-02942-42
2	75	52.1	25	5	PCT-US91-02942-26
3	60	41.7	13	3	US-08-783-853A-60
4	56	38.9	18	5	PCT-US95-06077-3
5	55	38.2	26	1	PCT-US95-06077-24
6	55	38.2	15	1	US-08-450-257-6
7	55	38.2	15	1	US-08-450-246-6
8	55	38.2	15	1	US-08-450-098-6
9	55	38.2	15	1	US-08-451-233-6
10	55	38.2	15	1	US-08-450-236-6
11	55	38.2	15	4	US-08-235-403-6
12	55	38.2	24	1	US-08-450-237-4
13	55	38.2	24	1	US-08-450-246-4
14	55	38.2	24	1	US-08-450-098-4
15	55	38.2	24	1	US-08-451-233-4
16	55	38.2	24	1	US-08-450-236-4
17	55	38.2	24	4	US-08-235-403-4
18	54	37.5	25	1	US-08-450-257-53
19	54	37.5	25	1	US-08-450-246-53
20	54	37.5	25	1	US-08-450-098-53
21	54	37.5	25	1	US-08-451-233-53
22	54	37.5	25	1	US-08-450-236-53
23	54	37.5	25	4	US-08-235-403-53
24	54	37.5	26	1	US-08-450-257-48
25	54	37.5	26	1	US-08-450-246-48
26	54	37.5	26	1	US-08-450-098-48
27	54	37.5	26	1	US-08-451-233-48

#### ALIGNMENTS

28	54	37.5	26	1	US-08-450-236-48	Sequence 48, Appl
29	54	37.5	26	4	US-08-235-403-48	Sequence 48, Appl
30	54	37.5	26	6	5169933-2	Patent No. 5169933
31	54	37.5	28	4	US-09-041-886-47	Sequence 47, Appl
32	54	37.5	29	2	US-08-928-958-9	Sequence 9, Appl
33	54	37.5	29	2	US-09-072-429-9	Sequence 9, Appl
34	54	37.5	29	4	US-09-411-706-5	Sequence 9, Appl
35	51	35.4	14	5	PCT-US95-06077-4	Sequence 4, Appl
36	51	35.4	14	5	PCT-US95-06077-10	Sequence 10, Appl
37	51	35.4	14	5	PCT-US95-06077-12	Sequence 12, Appl
38	51	35.4	28	4	US-09-041-886-40	Sequence 40, Appl
39	50.5	35.1	25	4	US-09-041-886-44	Sequence 44, Appl
40	50.5	35.1	28	4	US-09-041-886-37	Sequence 37, Appl
41	50.5	35.1	28	4	US-09-041-886-38	Sequence 38, Appl
42	50.5	35.1	28	4	US-09-041-886-39	Sequence 39, Appl
43	50.5	35.1	28	4	US-09-041-886-53	Sequence 53, Appl
44	50.5	35.1	28	4	US-09-041-886-54	Sequence 54, Appl
45	50	34.7	11	2	US-08-706-741B-54	Sequence 54, Appl

RESULT 1  
PCT-US91-02942-42  
Sequence 42, Application PC/TUS9102942

GENERAL INFORMATION:  
APPLICANT: ROTHLEIN, ROBERT

APPLICANT: ADAIR, JOHN R

APPLICANT: ARTHAL, DILBERT S

TITLE OF INVENTION: HUMANIZED CDR-GRAFTED ICAM-1 ANTIBODY

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:  
ADDRESS: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Ave. NW Suite 300  
CITY: Washington  
STATE: D.C.

COUNTRY: USA

ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/02942

FILING DATE: 19910429

CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9009549.8

FILING DATE: 27-APR-1990

ATTORNEY/AGENT INFORMATION:  
NAME: FOX, SAM L

REGISTRATION NUMBER: 30,353

REFERENCE/DOCKET NUMBER: 1011,0586600

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800

TELEFAX: (202) 833-8716

INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein

PCT-US91-02942-42

Query Match 54.9%, Score 79; DB 5; Length 25;

Best Local Similarity 88.9%; Pred. No. 4.2e-05;

Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 12 SSGGTFTLTITSSLOAED 29  
||||| :|

Db 2 SGSGTDFLTITSSLDQDD 19

RESULT 2  
PCT-US91-02942-26  
Sequence 26, Application PC/TUS9102942  
GENERAL INFORMATION:  
APPLICANT: ROTHEIN, ROBERT  
APPLICANT: ADAIR, JOHN R  
APPLICANT: ATHWAL, DILJEET S  
TITLE OF INVENTION: HUMANIZED CDR-GRAFTED ICAM-1 ANTIBODY  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Ave. NW Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/02942  
FILING DATE: 19910429  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9009549.8  
FILING DATE: 27-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: FOX, SAM L.  
REGISTRATION NUMBER: 30,353  
REFERENCE/DOCKET NUMBER: 1011.0586600  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US91-02942-26

Query Match 52.1%; Score 75; DB 5; Length 25;  
Best Local Similarity 83.3%; Pred. No. 0.00016;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 12 SGSGTDFLTITSSLDQDD 29  
Db 2 SGSGTDFLTITSSLDQDD 19

RESULT 3  
US-08-783-853A-60  
Sequence 60, Application US/08783853A  
Patent No. 6005091  
GENERAL INFORMATION:  
APPLICANT: Blackburn, Michael  
APPLICANT: Church, William  
APPLICANT: Gross, Mitchell  
APPLICANT: Feuerstein, Gloria  
APPLICANT: Nichols, Andrew  
APPLICANT: Padlan, Eduardo  
APPLICANT: Patel, Arundhal  
APPLICANT: Sylvester, Daniel  
TITLE OF INVENTION: ANTICOAGULANT AGENTS USEFUL IN TREATMENT  
TITLE OF INVENTION: OF THROMBOSIS  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/783,853A  
FILING DATE: 16-JAN-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/029,119  
FILING DATE: 24-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Baumeister, Kirk  
REGISTRATION NUMBER: 33,833  
REFERENCE/DOCKET NUMBER: P50438  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5096  
TELEFAX:

INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
US-08-783-853A-60

Query Match 41.7%; Score 60; DB 3; Length 13;  
Best Local Similarity 92.3%; Pred. No. 0.011;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 12 SGSGTDFLTITSS 24  
Db 1 SGSGTDFLTITSS 13

RESULT 4  
PCT-US95-06077-3  
Sequence 3, Application PC/TUS9506077  
GENERAL INFORMATION:  
APPLICANT: Immunobiology Research, Institute Inc.  
TITLE OF INVENTION: Vaccine Interdiction of Extracellular  
TITLE OF INVENTION: Transactivating Proteins of Human Immunodeficiency Virus  
TITLE OF INVENTION: And Other Chronically Infecting Viruses Employing Similar  
TITLE OF INVENTION: Intercellular Transactivating Strategies  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr, PO Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: PCT/US95/06077



Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 YEKRRRQRRRG 13  
1 YGRRRRQRRRG 13

RESULT 7  
US-08-450-246-6  
Sequence 6, Application US/08450246  
Patent No. 5670617  
GENERAL INFORMATION:  
APPLICANT: FRANKEL, Alan  
APPLICANT: PABO, Carl  
APPLICANT: BARSOUM, James G.  
APPLICANT: FAMELL, Stephen E.  
APPLICANT: PEPINSKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,246  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/235,403  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993  
APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-450-246-6

Query Match 38.2%; Score 55; DB 1; Length 15;  
Best Local Similarity 84.6%; Pred. No. 0.065;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 YEKRRRQRRRG 13  
1 YGRRRRQRRRG 13

DB 1 YGRRRRQRRRG 13

RESULT 8  
US-08-450-098-6  
Sequence 6, Application US/08450098  
Patent No. 5674980  
GENERAL INFORMATION:  
APPLICANT: FRANKEL, Alan  
APPLICANT: PABO, Carl  
APPLICANT: BARSOUM, James G.  
APPLICANT: FAMELL, Stephen E.  
APPLICANT: PEPINSKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,098  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/235,403  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993  
APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-450-098-6

Query Match 38.2%; Score 55; DB 1; Length 15;  
Best Local Similarity 84.6%; Pred. No. 0.065;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 YEKRRRQRRRG 13  
1 YGRRRRQRRRG 13  
RESULT 9

```

US-08-451/233-6
; Sequence 6, Application US/08451233
; Patent No. 5747641
; GENERAL INFORMATION:
; APPLICANT: FRANKEL, Alan
; APPLICANT: PABO, Carl
; APPLICANT: BARSOUM, James G.
; APPLICANT: FAWELL, Stephen E.
; APPLICANT: PEPINSKY, R. B.
; TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/451,233
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/235,403
; FILING DATE: 28-APR-1994
; APPLICATION NUMBER: US 07/934,375
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: US 07/098,766
; FILING DATE: 28-JUL-1993
; APPLICATION NUMBER: PCT/US93/07833
; FILING DATE: 19-AUG-1993
; APPLICATION NUMBER: US 07/454,450
; FILING DATE: 21-DEC-1989
; APPLICATION NUMBER: US 07/636,662
; FILING DATE: 02-JAN-1991
; APPLICATION NUMBER: US 08/158,015
; FILING DATE: 24-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: B170 CIP 2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-451-233-6

Query Match      38.2%; Score 55; DB 1; Length 15;
Best Local Similarity 84.6%; Pred. No. 0.065;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1 YEKRRRRRRSG 13
Db      1 YGRRRRRRRPG 13

RESULT 10
US-08-450-236-6
; Sequence 6, Application US/08450236
; Patent No. 5804604
; GENERAL INFORMATION:

```

```

; APPLICANT: FRANKEL, Alan
; APPLICANT: PABO, Carl
; APPLICANT: BARSOUM, James G.
; APPLICANT: FAWELL, Stephen E.
; APPLICANT: PEPINSKY, R. B.
; TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,236
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/235,403
; FILING DATE: 28-APR-1994
; APPLICATION NUMBER: US 07/934,375
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: US 07/098,766
; FILING DATE: 28-JUL-1993
; APPLICATION NUMBER: PCT/US93/07833
; FILING DATE: 19-AUG-1993
; APPLICATION NUMBER: US 07/454,450
; FILING DATE: 21-DEC-1989
; APPLICATION NUMBER: US 07/636,662
; FILING DATE: 02-JAN-1991
; APPLICATION NUMBER: US 08/158,015
; FILING DATE: 24-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: B170 CIP 2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-450-236-6

Query Match      38.2%; Score 55; DB 1; Length 15;
Best Local Similarity 84.6%; Pred. No. 0.065;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1 YEKRRRRRRSG 13
Db      1 YGRRRRRRRPG 13

RESULT 11
US-08-235-403-6
; Sequence 6, Application US/08235403
; Patent No. 6316003
; GENERAL INFORMATION:
; APPLICANT: FRANKEL, Alan
; APPLICANT: PABO, Carl
; APPLICANT: BARSOUM, James G.
; APPLICANT: FAWELL, Stephen E.

```

APPLICANT: PEPIISKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/235,403  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993  
APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-235-403-6

Query Match 38.2%; Score 55; DB 4; Length 15;  
Best Local Similarity 84.6%; Pred. No. 0.065;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YERKKRRRRRG 13  
Db 1 YGRKKRRRRRG 13

RESULT 12  
US-08-450-257-4  
Sequence 4, Application US/08450257  
Patent No. 5652122  
GENERAL INFORMATION:  
APPLICANT: FRANKEL, Alan  
APPLICANT: PABO, Carl  
APPLICANT: BARSDOM, James G.  
APPLICANT: FAMELL, Stephen E.  
APPLICANT: PEPIISKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020

CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,257  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/235,403  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993  
APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-450-257-4

Query Match 38.2%; Score 55; DB 1; Length 24;  
Best Local Similarity 84.6%; Pred. No. 0.11;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YERKKRRRRRG 13  
Db 10 YGRKKRRRRRG 22

RESULT 13  
US-08-450-246-4  
Sequence 4, Application US/08450246  
Patent No. 5670617  
GENERAL INFORMATION:  
APPLICANT: FRANKEL, Alan  
APPLICANT: PABO, Carl  
APPLICANT: BARSDOM, James G.  
APPLICANT: FAMELL, Stephen E.  
APPLICANT: PEPIISKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020



COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,246  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/235,403  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993  
APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-450-246-4

Query Match 38.2%; Score 55; DB 1; Length 24;  
Best Local Similarity 84.6%; Pred. No. 0.11;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YERKKRRRRSG 13  
Db 10 YGRKKRRRRRPG 22

RESULT 14  
US-08-450-098-4  
Sequence 4, Application US/08450098  
Patent No. 5674980  
GENERAL INFORMATION:  
APPLICANT: FRANKEL, Alan  
APPLICANT: PABO, Carl  
APPLICANT: BARSOUM, James G.  
APPLICANT: FAWELL, Stephen E.  
APPLICANT: PEPINSKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/451,233  
FILING DATE: 25-MAY-1995

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,098  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/235,403  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993  
APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-450-098-4

Query Match 38.2%; Score 55; DB 1; Length 24;  
Best Local Similarity 84.6%; Pred. No. 0.11;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YERKKRRRRSG 13  
Db 10 YGRKKRRRRRPG 22

RESULT 15  
US-08-451-233-4  
Sequence 4, Application US/08451233  
Patent No. 5747641  
GENERAL INFORMATION:  
APPLICANT: FRANKEL, Alan  
APPLICANT: PABO, Carl  
APPLICANT: BARSOUM, James G.  
APPLICANT: FAWELL, Stephen E.  
APPLICANT: PEPINSKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/451,233  
FILING DATE: 25-MAY-1995



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:52:41 ; Search time 25.22 Seconds

(without alignments)  
110.491 Million cell updates/sec

Title: US-09-712-819A-10

Perfect score: 144

Sequence: 1 YEKRRRRRRSSGGTATAPASSLQAEQ 29

Scoring table: BLOSUM62

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 5929

Minimum DB seq length: 0

Maximum DB seq length: 29

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	29.2	23	2	S43765 ribosomal protein
2	39	27.1	29	2	S08555 ribosomal protein
3	38	26.4	27	1	SRBPAS protamine A - stei
4	35	24.3	21	2	PN0882 sperm chromatin pr
5	35	24.3	25	1	JC4278 ribosomal protein
6	35	24.3	25	2	J01617 ribosomal protein
7	35	24.3	25	2	JC4685 ribosomal protein
8	35	24.3	25	2	S38425 ribosomal protein
9	35	24.3	25	2	T06233 ribosomal protein
10	35	24.3	25	2	T49214 ribosomal protein
11	34	23.6	25	2	A60741 insulin-like growt
12	34	23.6	27	1	SRAPC protamine B - Russ
13	33	22.9	15	2	A38304 heterogenous ribo
14	33	22.9	25	2	T38719 ribosomal protein
15	32	22.2	28	2	S26254 rel protein - chic
16	31	21.5	25	1	R6BY48 ribosomal protein
17	29	20.1	20	2	S09720 2S albumin small c
18	29	20.1	22	2	C64330 ribosomal protein
19	29	20.1	27	2	PC2316 multimeric end
20	29	20.1	29	2	S14040 hypothetical prote
21	28	19.4	10	2	PH0946 T-cell receptor be
22	28	19.4	12	2	A53524 ubiqlinol-cytochr
23	28	19.4	15	2	A49480 major immunophilin
24	28	19.4	16	2	A43552 caldesmon - turkey
25	28	19.4	19	2	A33361 CAMP-regulated pho
26	28	19.4	21	2	PC4443 cytochrome c3 - De
27	28	19.4	21	2	A36150 cytochrome c3 - De
28	28	19.4	23	2	A46186 ATP synthase beta
29	28	19.4	25	2	S07250 hypothetical prote

30	28	19.4	28	2	T37143 hypothetical prote
31	28	19.4	28	2	S68643 nicotinic acetylch
32	27.5	19.1	23	2	S47192 T-cell receptor J-
33	27	18.8	13	2	A53608 neurotensin - guin
34	27	18.8	17	2	E23734 insulin-like growt
35	27	18.8	18	2	PS0387 platelet-derived g
36	27	18.8	20	2	S38291 30K allergen - vel
37	27	18.8	23	2	S41390 p7 protein - human
38	27	18.8	27	2	S09504 homeotic protein c
39	27	18.8	29	2	TNLJBR trans-activating t
40	27	18.8	29	2	S35924 T-cell receptor ga
41	26.5	18.4	19	2	S12268 Oa-2 antigen - mou
42	26	18.1	10	2	E60787 sperm-activating p
43	26	18.1	13	1	UNBO neurotensin - bov
44	26	18.1	20	2	A61327 trypsin (EC 3.4.21
45	26	18.1	20	2	A25335 myosin-II light-chain

## ALIGNMENTS

RESULT 1  
S43765 ribosomal protein S19 - evening primrose mitochondrion (fragment)  
C:Species: mitochondrion Oenothera villaricae (evening primrose)  
C>Date: 10-Dec-1994 #sequence\_revision 12-May-1995 #text\_change 13-Aug-1999  
C:Accession: S43765; S30542  
R:Bock, H.; Brennicke, A.; Schuster, W.  
Plant Mol. Biol. 24, 811-818, 1994  
A>Title: Rps3 and rpl16 genes do not overlap in Oenothera mitochondria: GNG as a pote  
A:Reference number: S43765; MUID:94250844  
A:Accession: S43765  
A:Molecule type: DNA  
A:Residues: 1-23 <BOC>  
A:Cross-references: EMBL: X69140; NID: g13188; PIDN: CAA48893.1; PID: g13189  
A:Note: the source is designated as Oenothera berteriana  
C:Genetics:  
A:Gene: rps19  
A:Genome: mitochondrion  
C:Superfamily: Escherichia coli ribosomal protein S19  
C:Keywords: mitochondrion; protein biosynthesis; ribosome

Query Match 29.2%; Score 42; DB 2; Length 23;  
Best Local Similarity 53.3%; Pred. No. 41;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 1 YEKRRRRRRSSGSG 15  
DB 3 FTRKRRSRRTNIGSG 17

RESULT 2  
S08555 ribosomal protein L15 [validated] - Halobacterium salinarum (fragment)  
N:Alternate names: ribosomal protein HL16  
C:Species: Halobacterium salinarum  
C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 21-Jul-2000  
C:Accession: S08555  
R:Matheson, A.T.; Yaguchi, M.; Christensen, P.; Rollin, C.F.; Haasnaht, S.  
Can. J. Biochem. Cell Biol. 62, 426-433, 1984  
A>Title: Purification, properties, and N-terminal amino acid sequence of certain 50S  
A:Reference number: S07437; MUID:84282108  
A:Accession: S08555  
A:Molecule type: protein  
A:Residues: 1-29 <MAT>  
A:Note: the source is designated as Halobacterium cutirubrum  
A:Note: the protein is designated as ribosomal protein L16  
C:Superfamily: rat ribosomal protein L27a  
C:Keywords: protein biosynthesis; ribosome

Query Match 27.1%; Score 39; DB 2; Length 29;

Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 10; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

OY 4 KRRROR--RRSGSGT 16  
| | | | | | | | | |  
Db 3 KRRRGRGSRTHGCGT 17

## RESULT 3

SRAPAS  
protamine A - stellate sturgeon  
N:Alternate names: stellin  
C:Species: *Acipenser stellatus* (stellate sturgeon)  
C>Date: 28-Feb-1980 #sequence\_revision 12-Aug-1981 #text\_change 16-Feb-1997  
C:Accession: A02666  
R:Yulikov, E.P.; Rybin, V.K.; Silaev, A.B.  
Bioorg. Khim. 5, 5-10, 1979  
A:Title: The primary structure of stellin A.  
A:Reference number: A02666  
A:Accession: A02666  
A:Molecule type: protein  
A:Residues: 1-27 <YUL>  
A:Note: article in Russian with English abstract  
C:Superfamily: protamine Y2  
C:Keywords: chromosomal protein; DNA binding; spermatogenesis

Query Match 26.4%; Score 38; DB 1; Length 27;  
Best Local Similarity 70.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 4 KRRRORRRSG 13  
| - | | | | | | |  
Db 13 KRRRRRRRHHG 22

## RESULT 4

PN0082  
sperm chromatin protein 12-2 - Argentinian shortfin squid (fragment)  
C:Species: *Illex argentinus* (Argentinian shortfin squid)  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PN0082  
R:Osadchuk, L.A.; Levina, N.B.; Telezhinskaya, I.N.; Khrapunov, S.N.; Berdysher, G.D.; A  
Bioorg. Khim. 16, 448-455, 1990  
A:Title: Primary structure of main nuclear protein from headleg mollusk *Illex argentinus*  
A:Reference number: PN0081; MUID:90329035  
A:Accession: PN0082  
A:Molecule type: protein  
A:Residues: 1-21 <OSA>  
A:Experimental source: sperm  
C:Superfamily: sperm histone  
C:Keywords: sperm

Query Match 24.3%; Score 35; DB 2; Length 21;  
Best Local Similarity 58.3%; Pred. No. 3e+02;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 RKKRORRRSGS 14  
| : | | | | | | |  
Db 1 RRRRRRRRRRS 12

## RESULT 5

JC4278  
ribosomal protein L41, cytosolic [similarity] - rat  
C:Species: *Rattus norvegicus* (Norway rat)  
C>Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 01-Sep-2000  
C:Accession: JC4278  
R:Chan, Y.L.; Oliveira, J.; Wool, I.G.  
Biochem. Biophys. Res. Commun. 214, 810-818, 1995  
A:Title: The primary structures of rat ribosomal proteins L4 and L41.  
A:Reference number: JC4277; MUID:96024571

A:Accession: JC4278  
A:Molecule type: mRNA  
A:Residues: 1-25 <CHA>  
A:Cross-references: EMBL:X82550; NID:9575383; PIDN:CAA57899.1; PID:9575384  
C:Superfamily: rat ribosomal protein L41  
C:Keywords: ribosome

Query Match 24.3%; Score 35; DB 1; Length 25;  
Best Local Similarity 54.5%; Pred. No. 3.5e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 ERKKRORRRS 12  
| | | | | : | | |  
Db 14 KRRRRRRRRRS 24

## RESULT 6

JQ1617  
ribosomal protein L41.e - human  
N:Alternate names: HG12 protein; ribosomal protein YL41  
C:Species: *Homo sapiens* (man)  
C>Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 18-Aug-2000  
C:Accession: JQ1617; JCS659; S31691  
R:Klaudiny, J.; von der Kammer, H.; Scheit, K.H.  
Biochem. Biophys. Res. Commun. 187, 901-906, 1992  
A:Title: Characterization by cDNA cloning of the mRNA of a highly basic human protein  
A:Reference number: JQ1617; MUID:92412140  
A:Accession: JQ1617  
A:Molecule type: mRNA  
A:Residues: 1-25 <KLA>  
A:Cross-references: EMBL:Z12962; NID:936135; PIDN:CAA78306.1; PID:936136  
R:Lee, J.H.; Kim, J.M.; Kim, M.S.; Lee, Y.T.; Marshak, D.R.; Bae, Y.S.  
Biochem. Biophys. Res. Commun. 238, 462-467, 1997  
A:Title: The highly basic ribosomal protein L41 interacts with the beta subunit of pr  
A:Reference number: JCS659; MUID:97446005  
A:Accession: JCS659  
A:Molecule type: mRNA  
A:Residues: 1-25 <LEB>  
A:Cross-references: EMBL:Z12962; NID:936135; PIDN:CAA78306.1; PID:936136  
C:Comment: This protein stimulates phosphorylation of the beta chain of DNA topoisom  
C:Superfamily: rat ribosomal protein L41

Query Match 24.3%; Score 35; DB 2; Length 25;  
Best Local Similarity 54.5%; Pred. No. 3.5e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 ERKKRORRRS 12  
| | | | | : | | |  
Db 14 KRRRRRRRRRS 24

## RESULT 7

JC4685  
ribosomal protein L41 - cat  
C:Species: *Felis silvestris catus* (domestic cat)  
C>Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #text\_change 18-Aug-2000  
C:Accession: JC4685  
R:Starkey, C.R.; Menon, R.P.; Prabhu, S.; Levy, L.S.  
Biochem. Biophys. Res. Commun. 220, 648-652, 1996  
A:Title: Primary sequence and evolutionary conservation of ribosomal protein genes fr  
A:Reference number: JC4685; MUID:96183078  
A:Accession: JC4685  
A:Molecule type: mRNA  
A:Residues: 1-25 <STRA>  
A:Cross-references: GB:U22229; NID:9950108; PIDN:AA01667.1; PID:9950109  
A:Experimental source: thymic lymphoma  
C:Genetics:  
A:Gene: rPL41  
C:Superfamily: rat ribosomal protein L41

Query Match	24.38;	Score 35;	DB 2;	Length 25;
Best Local Similarity	54.58;	Pred. No. 3.5e+02;		
Matches	6;	Conservative	4;	Mismatches 1;
				Indels 0;
				Gaps 0;

```
QY      2 ERKKRRQRRRS 12
          :||:|:|:|
Db     14 KRKKRRKMRQRS 24
```

RESULT 8  
S38425  
Ribosomal protein GLA1 - upland cotton  
C:Species: Gossypium hirsutum (upland cotton)  
C:Date: 13-Jan-1995 #sequence  
C:Accession: S38425  
R:Turley, R.B.; Ferguson, D.L.; Meredith, W.R.  
submitted to the EMBL Data Library, October 1993  
A:Reference number: S38425  
A:Accession: S38425  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-25 <TUR>  
A:Cross-references: EMBL:X75423; NID:g407800; PIDN:CA53175.1; PID:g407801  
C:Superfamily: rat ribosomal protein L41

Query Match	24.38;	Score 35;	DB 2;	Length 25;
Best Local Similarity	54.58;	Pred. No. 3.5e+02;		
Matches	6; Conservative	4; Mismatches	1; Indels	0; Gaps 0;

```
QY 2 ERKKRRQRRRS 12
      :||:|:|:|
Db 14 KRKRKRMRQRS 24
```

RESULT 9  
T06233  
Ribosomal protein L41 - soybean (fragment)  
C:Species: Glycine max (soybean)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 18-Aug-2000  
C:Accession: T06233  
R:Mahalingam, R.; Knapp, H.T.  
submitted to the EMBL data library, February 1998  
A:Reference number: Z15555  
A:Accession: T06233  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-25 <MAH>  
A:Cross-references: EMBL:AF047051; NID:g2905777; PIDD:AA003557.1; PID:g2905778  
A:Experimental source: cultivar Essex; root  
C:Superfamily: rat ribosomal protein L41  
C:Keywords: protein biosynthesis; ribosome

Query Match	24.38;	Score 35;	DB 2;	Length 25;
Best Local Similarity	54.58;	Pred. NO. 3.5e+02;		
Matches	6; Conservative	4; Mismatches	1; Indels	0; Gaps 0;

```
QY      2 ERKKRRQRRS 12
          :||:|:|:|
Db      14 KKKRRKMRQS 24
```

RESULT 10  
T49214  
ribosomal protein GL41-like - Arabidopsis thaliana  
N:Alternate names: protein F7K19.200  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 18-Aug-2000  
C:Accession: T49214  
R:Benes, V., Wurmbach, E., Drzonek, H., Ansoyge, W., Mewes, H.W., Rudd, S., Lemcke, K.,  
submitted to the Protein Sequence Database, April 2000  
;Reference number: Z25014

A:Accession: T49214  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-25 <GEN>  
A:Cross-references: EMBL:AL163832; GSPDB:GN0061; ATSP:F27K19.2000  
A:Experimental source: cultivar Columbia, BAC clone F27K19  
C:Genetics:  
A:gene: ATSP:F27K19.200  
A:Map position: 3  
A:Introns: 4/5  
C:Superfamily: rat ribosomal protein L41

Query Match	24.38;	Score 35;	DB 2;	Length 25;
Best Local Similarity	54.5%;	Pred. No. 3.5e+02;		
Matches	6; Conservative	4; Mismatches	1; Indels	0; Gaps 0;

Oy	2	ERKKRRQRRRS	12
		:  : : :	
Db	14	KKKRRKMRQRS	24

```

RESULT 11
A60741
Insulin-like growth factor-binding protein Ie39L - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 28-Apr-1993 #sequence-revision 28-Apr-1993 #text-change 30-Sep-1993
C:Accession: A60741
R:Forbes, B.; Ballard, F.J.; Wallace, J.C.
J. Endocrinol. 126, 497-506, 1990
A:Title: An insulin-like growth factor-binding protein purified from medium condition
A:Reference number: A60741; MUID:91011238
A:Accession: A60741
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-25 <FOR>

```

Query Match	23.68;	Score 34;	DB 2;	Length 25;
Best Local Similarity	35.38;	Pred. No. 4.7e+02;		
Matches	6; Conservative	2; Mismatches	9; Indels	0; Gaps 0;

```
QY      13 GSGTDATAPASSLOAED 29
          | | | | : : | |
Db      6 GQGVQAGXPGGXEVED 22
```

RESULT 12  
SRAPC  
protamine B - Russian sturgeon  
N/Alternate names: sturine  
C/Species: Acipenser gueldenstaedti (Russian sturgeon)  
C/Date: 12-Aug-1981 #sequence\_revision 12-Aug-1981 #text\_change 16-Feb-1997  
C/Accession: A02665  
R/Yulikova, E.P.; Evsenko, L.K.; Baratova, L.A.; Belyanova, L.P.; Rydin, V.K.; Silaev  
B/Orig. Khlm. 2, 1613-1617, 1976  
A/Reference number: A02665  
A/Accession: A02665  
A/Molecule type: protein  
A/Residues: 1-27 <YUL>  
C/Superfamily: protamine Y2  
;Keywords: chromosomal protein; DNA binding; spermatogenesis

Query Match	23.68;	Score 34;	DB 1;	Length 27;
Best Local Similarity	60.0%;	Pred. No. 5e+02;		
Matches	6; Conservative	3; Mismatches	1; Indels	0; Gaps 0;

QY	4	KRRQRSSG	13
		::  :	1
Db	11	QRRRRRRHG	20

RESULT 13  
A38304  
heterogeneous ribonuclear particle protein A1 - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 14-Jun-1991 #sequence\_revision 14-Jun-1991 #text\_change 31-Oct-1997  
C:Accession: A38304  
R:Kumar, A.; Casas-Finef, J.R.; Luneau, C.J.; Karpel, R.L.; Merrill, B.M.; Williams, K.F.  
J. Biol. Chem. 265, 17094-17100, 1990  
A:Title: Mammalian heterogeneous nuclear ribonucleoprotein A1. Nucleic acid binding prop  
A:Reference number: A38304; MUID:91009136  
A:Accession: A38304  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-15 <KUN>  
C:Superfamily: helix-destabilizing protein; ribonucleoprotein repeat homology

Query Match 22.9%; Score 33; DB 2; Length 15;  
Best Local Similarity 87.5%; Pred. No. 3.9e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 ORRRSGSG 15  
|||  
DB 8 ORRRSGSG 15

RESULT 14  
T38719  
ribosomal protein L41 - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Aug-2000  
C:Accession: T38719; T38731; T39176  
R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsby, S.V.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: 221807  
A:Accession: T38719  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-25 <MUR>  
A:Cross-references: EMBL:Z69369; PIDN:CAB40187.1; GSPDB:GN00066  
A:Experimental source: strain 972h-; cosmid c3F10  
R:Gentles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: 221797  
A:Accession: T38731  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-25 <GEN>  
A:Cross-references: EMBL:Z29167; PIDN:CAB40152.1; GSPDB:GN00066; SPDB:SPAC3G6.13C  
A:Experimental source: strain 972h-; cosmid c3G6  
R:Bothe, G.; Pohl, T.; McDougall, R.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, August 1999  
A:Reference number: 221833  
A:Accession: T39176  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-25 <BOT>  
A:Cross-references: EMBL:AL109738; PIDN:CAB52162.1; GSPDB:GN00066; SPDB:SPAC8F11.01C  
A:Experimental source: strain 972h-; cosmid c8F11  
C:Genetics:  
A:Gene: SPAC31F10.18C; SPAC3G6.13C; SPAC31F10.18C; SPAC8F11.01C  
A:Map position: 1  
A:Introns: 24/2  
C:Superfamily: rat ribosomal protein L41

Query Match 22.9%; Score 33; DB 2; Length 25;  
Best Local Similarity 54.5%; Pred. No. 6.3e+02;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 LKRRRORRRS 12  
:|:|:|:|  
DB 14 KRRRYMRARS 24

RESULT 15  
S26254  
rel protein - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 25-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 06-Dec-1996  
C:Accession: S26254  
R:Capobianco, A.J.; Gilmore, T.D.  
Oncogene 6, 2203-2210, 1991  
A:Title: Repression of the chicken c-rel promoter by vRel in chicken embryo fibroblas  
A:Reference number: S26254; MUID:92115319  
A:Accession: S26254  
A:Status: preliminary; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-28 <CAP>  
A:Cross-references: EMBL:X59588

Query Match 22.2%; Score 32; DB 2; Length 28;  
Best Local Similarity 36.4%; Pred. No. 9.3e+02;  
Matches 8; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 YERKKRORRRSGGTDPAPA 22  
:|:|:|:|:|:|  
DB 5 FVRGAARSPRRGSPSPDAELSA 26

Search completed: July 15, 2002, 13:52:41  
Job time: 160 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:57:02 ; Search time 12.84 Seconds  
(without alignments)  
87.451 Million cell updates/sec

Title: US-09-712-819a-10  
Perfect score: 144  
Sequence: 1 YERKKRQRGRSGGTATAASSLQAE D 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 1766

Minimum DB seq length: 0  
Maximum DB seq length: 29

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	39	27.1	29	RL15_HALCU	P05971 halobacteri
2	38	26.4	27	PRTA_ACIST	P02324 acipenser s
3	35	24.3	25	RL11_ARATH	P35015 arabidopsis
4	35	24.3	25	RL11_HUMAN	P28751 homo sapien
5	35	24.3	25	RL11_SOYBN	O49224 glycine max
6	34	23.6	27	PRTB_ACIGU	P02323 acipenser g
7	33	22.9	25	RL11_SCHPO	O9Y710 schizosacch
8	32	22.2	25	RL11_AGABI	P78569 agaricus bi
9	32	22.2	25	RL11_QUESU	O62713 quercus sub
10	31	21.5	25	RL11_YEAST	P05746 saccharomyc
11	29	20.1	22	RL11_METJA	P54025 methanococc
12	28	19.4	29	TL16_SPTOL	P21834 spinachia ol
13	27	18.8	13	NEOT_CAVPO	P32560 cavia porce
14	27	18.8	15	UC06_MAIZE	P80612 zea mays (m
15	27	18.8	27	HM12_CAFEL	P17487 caenorhabdi
16	25	17.4	27	LIFE_BOVIN	P16386 bos taurus
17	24	16.7	19	ANP7_ELEGR	P11920 eleginus gr
18	24	16.7	19	FIRA_CPRET	P14446 cervus elap
19	24	16.7	19	FIRA_CPRET	P14447 cervus elap
20	23	16.0	15	ATP2_SPTOL	P80083 spinachia ol
21	23	16.0	15	UC30_MAIZE	P80636 zea mays (m
22	23	16.0	24	VG1_BPAL3	P08766 bacterioph
23	23	16.0	25	RS11_THERH	P80372 theirus aqu
24	23	16.0	27	PPS1_DROFU	P01372 drosophila
25	23	16.0	28	VG9_SPV4	P11341 spiroplasma
26	22	15.3	12	UR2A_CATCO	P04558 catostomus
27	22	15.3	18	YMOB_CHIAT	O45827 chloroflexu
28	22	15.3	19	PSAF_CUCSA	P42047 cucumis sat
29	22	15.3	22	RL18_HALME	P30561 halobacteri
30	22	15.3	23	RL18_HALVO	P50563 halobacteri
31	22	15.3	24	KAD_BACLI	P35140 bacillus li
32	22	15.3	26	RL18_HALMA	P50566 halobacteri
33	21.5	14.9	28	CH60_MYCSM	P80673 mycobacteri

34	21	14.6	15	1	K10M_LIMTE
35	21	14.6	16	1	R1PK_TTRKI
36	21	14.6	18	1	YAA5_RHOPA
37	21	14.6	20	1	ATP4_SPTOL
38	21	14.6	20	1	PSAK_PEA
39	21	14.6	21	1	NRLA_ACISP
40	21	14.6	22	1	IF2G_PIG
41	21	14.6	22	1	MOT1_CANFA
42	21	14.6	24	1	CSMB_PELIU
43	21	14.6	25	1	ANDT_ANDAU
44	21	14.6	25	1	VCJ_BPG4
45	21	14.6	29	1	ND01_SQTLU

## ALIGNMENTS

RESULT 1  
ID RL15\_HALCU STANDARD; PRT; 29 AA.  
AC P05971;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE 50S ribosomal protein L15P (HL16) (Fragment).  
GN RPL15P.  
OS Halobacterium cutirubrum.  
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
OC Halobacterium.  
OX NCBI\_TaxID=2242;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=84282108; PubMed=6467081;  
RA Matheson A.T., Yaguchi M., Christensen P., Rollin C.F., Hasnain S.;  
RT "Purification, properties, and N-terminal amino acid sequence of  
RT certain 50S ribosomal subunit proteins from the archaeobacterium  
RT Halobacterium cutirubrum.";  
RL Can. J. Biochem. Cell Biol. 62:426-433(1984).  
CC -1 SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.  
DR PIR: S08555; S08555.  
DR InterPro: IPR001196; Ribosomal L15.  
DR Prosite: P500475; RIBOSOMAL\_L15; PARTIAL.  
KW Ribosomal protein.  
FT NON\_TER 29  
SQ SEQUENCE 29 AA: 3170 MW: 73D13D831B4BE790 CRC64:

Query Match 27.1%; Score 39; DB 1; Length 29;  
Best local Similarity 66.7%; Pred. No. 29;  
Matches 10; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

OY 4 KKRQR--RRSGGT 16  
DB 3 KKRQRGRSGGT 17

RESULT 2  
ID PRTA\_ACIST STANDARD; PRT; 27 AA.  
AC P02324;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Proteolines A and C (Stellin A and C).  
OS Acipenser stellatus (Sevruga).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Chondrostei; Acipenseriformes; Acipenseridae;  
OC Acipenseridae; Acipenser.  
OX NCBI\_TaxID=7903;  
RN [1]  
RP SEQUENCE (STELLIN A).  
RA Yulikova E.P., Rybin V.K., Silaev A.B.;  
RT "The primary structure of stellin A.";

RL Bloorg. Khlm. 5:5-10(1979).  
RN [12]  
RP SEQUENCE (STELLIN C).  
RX MEDLINE-94019256; PubMed-1339950;  
RA Rybin V.K., Revina L.P., Bartova L.A.;  
RT "Comparison of amino acid sequences of sturgeon triptocamines using  
RL protamines from *Acipenser stellatus* gonads as an example.";  
Mol. Biol. (Mosk) 26:300-306(1992).  
CC -1- FUNCTION: PROTAGENINS SUBSTITUTE FOR HISTONES IN THE CHROMATIN OF  
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT  
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: TESTIS.  
DR PIR: A02666; SRAPAS.  
KM Chromosomal protein, Nucleosome core; Spermatogenesis; DNA-binding;  
KM Testis; DNA condensation; Nuclear protein.  
FT VARIANT 1 MISSING (IN STELLIN C).  
SQ SEQUENCE 27 AA: 3532 MW: 3EAA45B950FA8658 CRC64;  
  
Query Match 26 4%; Score 38; DB 1; length 27;  
Best Local similarity 70.0%; Pired. No. 37;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 4 KRRRRRRSG 13  
DB 13 KRRRRRRRHG 22  
  
RESULT 3  
RL41\_ARATH STANDARD: PRT: 25 AA.  
AC P35015;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE 60S ribosomal protein L41.  
GN (RP41-A OR AT1G56045 OR T6H22.15) AND (RP41-B OR AT3G08520 OR  
GN T6G24.5 OR T6G24.5) AND (RP41-C OR AT3G11120 OR F9F8.7) AND  
GN (RP41-D OR AT3G56020 OR F2K19.200)  
OS Arabidopsis thaliana (Mouse-ear cress).  
OS Nicotiana tabacum (Common tobacco).  
OS Gossypium hirsutum (Upland cotton).  
OS Pisum sativum (Garden pea).  
OS Hordeum vulgare (Barley), and  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702; 4097, 3635, 3888, 4513, 4530;  
RN [1]  
RP SEQUENCE FROM N.A. (AT1G56045).  
RC SPECIES-A.thaliana; STRAIN-CV. COLUMBIA;  
RX MEDLINE-21016719; PubMed-1130712;  
RA Theologos A., Ecker J.R., Palm C.J., Federpriel N.A., Kaul S.,  
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,  
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,  
RA Dunn P., Elgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,  
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maltl R., Marziani A.,  
RA Miltcher J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I.,  
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
RA Sakano H., Salberg S.L., Schwartz J.R., Slinn P., Southwick A.M.,  
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
RA Utechtack T., Van Aken S., Vayenberg M., Vayotskaya V.S., Walker M.,  
RA Wu L., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
thaliana.";  
RL Nature 408:816-820(2000).  
  
RN [12]  
RP SEQUENCE FROM N.A. (AT3G08520; AT3G11120 AND AT3G56020).  
RC SPECIES-A.thaliana; STRAIN-CV. COLUMBIA;  
RX MEDLINE-21016720; PubMed-1130713;  
RA Salanoubat M., Lemcke K., Rieger M., Ansoorge W., Unselid M.,  
RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermayer B.,  
RA Deleney M., Boutry M., Grievell L.A., Macie R., Pulgomech P.,  
RA De Simone V., Choise N., Artiguenave F., Robert C., Broillet F.,  
RA Wincker P., Catolico L., Weissmann J., Saurin W., Queller F.,  
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
RA Wummbach E., Dronok H., Erle H., Jordan N., Bangerl S.,  
RA Wiedemann R., Kranz H., Voss H., Holland N., Brandt P., Nyakatura G.,  
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonetti B.,  
RA Conrad A., Hornischer K., Kauer G., Loehmet T.-H., Nordstet B.,  
RA Reichelt J., Scharfe M., Schoen O., Barques M., Terol J., Clement J.,  
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,  
RA Cooke R., Landie M., Berger-Llauro C., Purrelle B., Masny D.,  
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cortet A., Cascuberta E.,  
RA Monfort A., Argilou A., Flores M., Ligouri R., Vitale D.,  
RA Mannhapt G., Haase D., Schoof H., Rud S., Zaccaria P., Bewes H.-W.,  
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
RA Rooney T., Rizzo M., Walts A., Utechtack T., Fujii C.Y., Shea T.P.,  
RA Cressy T.H., Haas B., Maltl R., Wu D., Peterson J., Van Aken S.,  
RA Pal G., Miltcher J., Sellers P., Gill J.E., Feldblyum T.V.,  
RA Preuss D., Lin X., Niernan W.C., Salberg S.L., White O., Venter J.C.,  
RA Fraser C.M., Kaneko T., Idesawa K., Kawashima K., Kishida Y.,  
RA Sasamoto S., Kimura T., Nakamura Y., Sato S., Kato T., Asamizu E.,  
RA Kiyokawa G., Kohara M., Matsumoto H., Matsuno A., Muraki A.,  
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,  
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;  
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
thaliana.";  
RL Nature 408:820-822(2000).  
RN [13]  
RP SEQUENCE FROM N.A.  
RC SPECIES-A.thaliana; STRAIN-CV. COLUMBIA;  
RX Chouk R., Chen H., Kim C.J., Koeseema E., Meyers M.C., Banh J.,  
RA Bower J., Carlini P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
RA Ishida J., Jiang P.X., Jones T., Kamlay A.E., Karlin-Neumann G.,  
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M.,  
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sekuraki T.,  
RA Sacou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
RA Ecker J.R.;  
RT "Arabidopsis cDNA clones";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
RN [14]  
RP SEQUENCE FROM N.A.  
RC SPECIES-N.tabacum; STRAIN-CV. SRI; TISSUE=Leaf;  
RX Zhou X.R.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
RN [15]  
RP SEQUENCE FROM N.A.  
RC SPECIES-G.hirsutum; STRAIN-CV. DELTAPEINE 62;  
RX MEDLINE-93062728; PubMed-7972506;  
RA Tutley R.B., Ferguson D.L., Meredith W.R.;  
RT "Isolation and characterization of a cDNA encoding ribosomal protein  
L41 from cotton (*Gossypium hirsutum* L.)";  
RL Plant Physiol. 105:1449-1450(1994).  
RN [16]  
RP SEQUENCE FROM N.A.  
RC SPECIES-P.sativum; STRAIN-CV. LITTLE MARVEL; TISSUE=Root tip;  
RA Woo H.H.;  
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
RN [17]  
RP SEQUENCE FROM N.A.  
RC SPECIES-H.vulgare; STRAIN-CV. Boml; TISSUE=Endosperm;  
RX Rasmussen S.K.;  
RT "Barley L41 ribosomal protein from immature endosperm.";  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
RN [18]  
RP SEQUENCE FROM N.A.  
RC SPECIES-O.sativa; STRAIN-CV. Nipponbare;



RA Buell C.R., Yuan O., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,  
 RA Zismann V., Pai G., Bowman C.L., Fujii C.Y., Vanaken S.E.,  
 RA Bowman C.L., Craven B., Utterback T.R., Khakh H., Feldtlyum T.V.,  
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.,  
 RT "Oryza sativa chromosome 3 BAC OSJNBa0091j19 genomic sequence."  
 CC Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 CC - SIMILARITY: BELONGS TO THE LAIE FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC DR EMBL: AC009894; AAF02845.1; -  
 DR EMBL: AC074395; AAC50829.1; -  
 DR EMBL: AC009991; AAF01511.1; -  
 DR EMBL: AL163832; CAB87856.1; -  
 DR EMBL: AY058053; AAL24161.1; -  
 DR EMBL: U26255; AAA67297.1; -  
 DR EMBL: X75423; CAA53175.1; -  
 DR EMBL: L47967; AAA79268.1; -  
 DR EMBL: AJ001160; CAA04564.1; -  
 DR EMBL: AC084320; AAK09215.1; -  
 DR PIR: S38425; S38425.  
 KM Ribosomal protein.  
 SQ SEQUENCE 25 AA; 3428 MW; 9AD629DA293C039E CRC64;

Query Match 24.3%; Score 35; DB 1; Length 25;  
 Best Local Similarity 54.5%; Pred. No. 84;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 ERKRRRRRS 12  
 :1:1: 1:11  
 Db 14 KRRRRRRRS 24

RESULT 4  
 RI41\_HUMAN STANDARD; PRT; 25 AA.  
 AC P28731.  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE 60S ribosomal protein L41 (HG12).  
 GN RPL41.  
 OS Homo sapiens (Human).  
 OS Mus musculus (Mouse).  
 OS Rattus norvegicus (Rat).  
 OS Felis silvestris catus (Cat).  
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxId=9606, 10090, 10116, 9685, 31033;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Human; TISSUE-Ovary.  
 RX MEDLINE=92412140; PubMed=136959;  
 RA Klaidun J., von der Kammer H., Scheit R.H.,  
 RT "Characterization by cDNA cloning of the mRNA of a highly basic human  
 RT protein homologous to the yeast ribosomal protein YL41.";  
 RL Biochem. Biophys. Res. Commun. 187:901-906(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Human;  
 RX MEDLINE=97446005; PubMed=9299532;  
 RA Lee J.-H., Kim J.-M., Kim M.-S., Lee Y.-T., Marshak D.R., Bae Y.-S.,  
 RT "The highly basic ribosomal protein L41 interacts with the beta  
 RT subunit of protein kinase CKII and stimulates phosphorylation of DNA  
 RT topoisomerase Ialpha by CKII.";

RL Biochem. Biophys. Res. Commun. 238:462-467(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Human;  
 RA Go H., Miyado K., Hasegawa H., Taniguchi S.,  
 RT "Characterization of human ribosomal protein L41 genomic structure."  
 RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Human;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burdill W., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R., Corry N.R.,  
 RA Coulson A., Coville G.J., Deacon R., Dhami P., Dunn H.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jakosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Lehelvalho M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConnachie L.J., McEay K., McMurtry A.A.,  
 RA Milne S., Misty D., Moore M.D.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Beck A.L.,  
 RA Phillimore B.J.C.T., Prithalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Seltra H.K., Showkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.,  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Rat; STRAIN-SPRAGUE-DAWLEY; TISSUE-Liver;  
 RA Chan Y.-L., Wool T.G.,  
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Mouse; STRAIN-C57BL/6;  
 RA Rocha D., Anderson E., Botcherby M., Jordan B.,  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-F.s.catus;  
 RX MEDLINE=96183078; PubMed=8607819;  
 RA Starkey C.R., Menon R.P., Prabhur S., Levy L.S.,  
 RT "Primary sequence and evolutionary conservation of ribosomal protein  
 RT genes from the domestic cat.";  
 RL Biochem. Biophys. Res. Commun. 220:648-652(1996).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-F.rubripes;  
 RX MEDLINE=9177347; PubMed=10077531;  
 RA Gellner K., Brenner S.,  
 RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu  
 RT rubripes.";  
 RL Genome Res. 9:251-258(1999).  
 CC -1- FUNCTION: INTERACTS WITH THE BETA SUBUNIT OF PROTEIN KINASE CKII  
 CC AND STIMULATES PHOSPHORYLATION OF DNA TOPOISOMERASE II ALPHA BY  
 CC CKII.  
 CC -1- SIMILARITY: BELONGS TO THE LAIE FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

```

CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: Z12862; CAA78306.1; -.
DR  EMBL: AF026844; AAB82715.1; -.
DR  EMBL: AB010874; BAA31508.1; -.
DR  EMBL: AL035562; CAB46994.1; -.
DR  EMBL: X82550; CAA57899.1; -.
DR  EMBL: U93862; AAB52254.1; -.
DR  EMBL: U22229; AAB01667.1; -.
DR  EMBL: AF056116; AAC34393.1; -.
DR  PIR: JQ1617; JQ1617.
DR  PIR: S31691; S31691.
KW  Ribosomal protein.
SQ  SEQUENCE 25 AA; 3456 MW; 9AD629C69ECC039E CRC64;

Query Match      24.3%; Score 35; DB 1; Length 25;
Best Local Similarity 54.5%; Pred. No. 84;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY  2 ERKRRRRRRS 12
    :|:|:|:|:|
DB  14 KRRKKRRMORS 24

RESULT  5
RL41_SOYBN      STANDARD;      PRT;      25 AA.
ID  RL41_SOYBN
AC  O49224;
DT  15-DEC-1998 (Rel. 37, Created)
DT  15-DEC-1998 (Rel. 37, Last sequence update)
DT  15-DEC-1998 (Rel. 37, Last annotation update)
DE  60S ribosomal protein L41.
GN  RPL41.
OS  Glycine max (Soybean).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC  eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX  NCBI_TaxID=3847;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CV, ESSEX; TISSUE=Root;
RL  Moe H.-H.;
RL  Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC  -1- SIMILARITY: BELONGS TO THE LALE FAMILY OF RIBOSOMAL PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  DR  EMBL: AF047051; AAC03557.1; -.
CC  KW  Ribosomal protein.
SQ  SEQUENCE 25 AA; 3130 MW; AD933FB8E49EDD6E CRC64;

Query Match      24.3%; Score 35; DB 1; Length 25;
Best Local Similarity 54.5%; Pred. No. 84;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY  2 ERKRRRRRRS 12
    :|:|:|:|:|
DB  14 KRRKKRRMORS 24

RESULT  5
PRTB_ACIGU      STANDARD;      PRT;      27 AA.
ID  PRTB_ACIGU
AC  P03323; ,

```

```

DT  21-JUL-1986 (Rel. 01, Created)
DT  21-JUL-1986 (Rel. 01, Last sequence update)
DT  01-DEC-1992 (Rel. 24, Last annotation update)
DE  Protamine B (Sturine B) (Stellin B).
OS  Acipenser guldenstadti (Caspian sturgeon) (Russian sturgeon), and
OS  Acipenser stellatus (Sevruga).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Actinopterygii; Chondrostei; Acipenseriformes; Acipenseridae;
OC  Acipenserinae; Acipenser.
OX  NCBI_TaxID=7902, 7903;
RN  [1]
RP  SEQUENCE.
RC  SPECIES=A.guldenstadti;
RA  Yulikova E.P., Evsenko L.K., Baratova L.A., Belyanova L.P.,
RA  Rybin V.K., Silaev A.B.;
RT  "The primary structure of sturine B, a protamine from Caspian
RT  sturgeon.";
RL  Bioorg. Khim. 2:1613-1617(1976).
RN  [2]
RP  SEQUENCE.
RC  SPECIES=A.stellatus;
RA  Rybin V.K., Yulikova E.P.;
RL  Khim. Prirod. Soedin. 5:700-704(1979).
CC  -1- FUNCTION: PROTAMINES SUBSTITUTE FOR HISTONES IN THE CHROMATIN OF
CC  SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC  SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC  -1- SUBCELLULAR LOCATION: Nuclear.
CC  -1- TISSUE SPECIFICITY: TESTIS.
DR  PIR: A02665; SRAPC.
KW  Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW  Testis; DNA condensation; Nuclear protein.
SQ  SEQUENCE 27 AA; 3707 MW; E30DB46B1737EC80 CRC64;

Query Match      23.6%; Score 34; DB 1; Length 27;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY  4 KRRRRRRSG 13
    :|:|:|:|:|
DB  11 ORRRRRRRHG 20

RESULT  7
RL41_SCHPO      STANDARD;      PRT;      25 AA.
ID  RL41_SCHPO
AC  Q9Y710;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  60S ribosomal protein L41.
GN  RPL41A OR SPAC36.13C AND (RPL41B OR SPAC31P10.18C OR SPAC3F11.01C).
OS  Schizosaccharomyces pombe (Fission yeast).
OC  Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC  Schizosaccharomycetales; Schizosaccharomycetaceae;
OC  Schizosaccharomyces.
OX  NCBI_TaxID=4896;
RN  [1]
RP  SEQUENCE FROM N.A. (RPL41A).
RC  STRAIN=972;
RA  Gentles S., Churcher C.M., Barrrell B.G., Rajandream M.A., Wood V.;
RL  Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A. (RPL41B).
RC  STRAIN=972;
RA  Bothe G., Pohl T., McDougall R., Rajandream M.A., Barrrell B.G.;
RL  Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC  -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L41 IN S.POMBE.
CC  -1- SIMILARITY: BELONGS TO THE LALE FAMILY OF RIBOSOMAL PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its

```

```

CC user by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; Z99167; CAB40152.1; -.
CC DR EMBL; Z69369; CAB40187.1; -.
CC DR EMBL; AL109738; CAB52162.1; -.
CC DR Ribosomal protein; Multigene family.
CC KW SEQUENCE 25 AA; 3411 MW; 9EC629DD9ECC0535 CRC64;
CC -----
Query Match 22.9%; Score 33; DB 1; Length 25;
Best Local Similarity 54.5%; Pred. No. 1.5e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
OY 2 ERKRRRRRRS 12
   :|:|:|:|
Db 14 KRRKRRRRARS 24
-----
RESULT 8
RL41_AGABI STANDARD; PRT; 25 AA.
ID P78569;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 60S ribosomal protein L41.
GN RL41.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HORST U1;
RA MEDLINE=97111990; PubMed=8953726;
RA Sonenberg A.S.M., de Groot P.W.J., Schaap P.J., Baars J.J.P.,
RA Visser J., van Gilsenven L.J.L.D.;
RT "Isolation of expressed sequence tags of Agaricus bisporus and their
RT assignment to chromosomes."
RL Appl. Environ. Microbiol. 62:4542-4547(1996)
CC -I- SIMILARITY: BELONGS TO THE L41E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X94764; CA64390.1; -.
CC DR Ribosomal protein.
CC KW SEQUENCE 25 AA; 3413 MW; A97629D0143C04EA CRC64;
CC -----
Query Match 22.2%; Score 32; DB 1; Length 25;
Best Local Similarity 60.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
OY 3 RRRRRRRRS 12
   :|:|:|:|
Db 15 RRRRRRRARS 24
-----
RESULT 9
RL41_QUEST STANDARD; PRT; 25 AA.
ID 082713;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

```

```

DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L41.
GN RPL41.
OS Quercus suber (Cork oak).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosids I; Fagales; Fagaceae; Quercus.
OX NCBI_TaxID=58331;
RN [1]
RP SEQUENCE FROM N.A.
RA Huguet G., Pla M., Verdaguer D., Molinas M.;
RT "Ribosomal proteins in Quercus suber."
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: BELONGS TO THE L41E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AJ001347; CA04691.1; -.
CC DR Ribosomal protein.
CC KW SEQUENCE 25 AA; 3337 MW; 6BC629DD9EDB3383 CRC64;
CC -----
Query Match 22.2%; Score 32; DB 1; Length 25;
Best Local Similarity 54.5%; Pred. No. 2.1e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
OY 2 ERKRRRRRRS 12
   :|:|:|:|
Db 14 KRRKRRRRARS 24
-----
RESULT 10
RL41_YEAST STANDARD; PRT; 25 AA.
ID P05746;
AC 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L41 (Yl41) (Yl47).
GN (RPL41A OR RPL47A OR YL41A OR YDL184C OR D1290) AND
GN (RPL41B OR RPL47B OR YL41B OR YDL133C).
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A. (RPL41A AND RPL41B).
RC MEDLINE=90234826; PubMed=2187623;
RA Suzuki K., Hashimoto T., Otake E.;
RT "Yeast ribosomal proteins: XI. Molecular analysis of two genes
RT encoding Yl41, an extremely small and basic ribosomal protein, from
RT Saccharomyces cerevisiae."
RL Curr. Genet. 17:185-190(1990).
RN [2]
RP SEQUENCE FROM N.A. (RPL41A).
RC STRAIN=S288C / FY1679;
RX MEDLINE=96021607; PubMed=8533471;
RA Vernasselt P., Voet M., Volckaert G.;
RT "New open reading frames, one of which is similar to the n1v gene of
RT Azotobacter vinelandii, found on a 12.5 kbp fragment of chromosome IV
RT of Saccharomyces cerevisiae."
RL Yeast 11:961-966(1995).
RN [3]
RP SEQUENCE FROM N.A. (RPL41B).
RC STRAIN=S288C / FY1679;
RX MEDLINE=97127826; PubMed=8972577;
RA Woelfl S., Haneman V., Saluz H.P.;

```

```

RT      *Analysis of a 26,756 bp segment from the left arm of yeast
RL      chromosome IV.
RN      Yeast 12:1549-1554(1996).
RP      SEQUENCE.
RA      Otsuka E., Higo K.-I., Itoh T.
RT      *Yeast ribosomal proteins: VIII. Isolation of two proteins and
RT      sequence characterization of twenty-four proteins from cytoplasmic
RT      ribosomes.
RL      Mol. Gen. Genet. 195:544-546(1984).
CC      -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L41 IN YEAST.
CC      -1- SIMILARITY: BELONGS TO THE L41E FAMILY OF RIBOSOMAL PROTEINS.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X16065; CAA34201.1; -
DR      EMBL; X16066; CAA34202.1; -
DR      EMBL; X83276; CAA58262.1; -
DR      EMBL; Z74232; CAA98759.1; -
DR      EMBL; Z74181; CAA98705.1; -
DR      EMBL; Z74182; CAA98706.1; -
DR      EMBL; X96876; CAA65626.1; -
DR      PIR; S22247; R6BY4A.
DR      PIR; S22246; R6BY4B.
DR      SGD; S0002343; RPL1A.
DR      SGD; S0002293; RPL1B.
KW      Ribosomal protein; Multigene family.
SQ      SEQUENCE 25 AA: 3337 MW; BD2629D9EDB5381 CRC64;

Query Match          21.5%; Score 31; DB 1; Length 25;
Best Local Similarity 54.5%; Pred. No. 2.8e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY      2 ERKRRORRRS 12
DB      14 RRRRRVARS 24
      :|:|:|:|
      :|:|:|:|

RESULT 11
RL1_METUA STANDARD: PRT; 22 AA.
AC      P34025;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      50S ribosomal protein L41E.
GN      RPL41E OR MJO242.
OS      Methanococcus jannaschii.
OC      Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC      Methanococcus.
OX      NCBI_TaxID=2190;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX      MEDLINE=96337999; Pubmed=8688087;
RA      Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA      Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA      Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Relch C.I.,
RA      Overbeek R., Kirkness E.F., Weinslock K.G., Merrick J.M., Giledek A.,
RA      Scott J.L., Geoghegan N.S.M., Weisman J.E., Puhmann J.L., Nguyen D.,
RA      Uuterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA      Cotton M.D., Roberts K.M., Hurst M.A., Kalne B.P., Borodovsky M.,
RA      Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA      *Complete genome sequence of the methanogenic archaeon, Methanococcus
RA      jannaschii.
RT      Science 273:1058-1073(1996).
RL

```

```

CC      -1- SIMILARITY: BELONGS TO THE L41E FAMILY OF RIBOSOMAL PROTEINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U67480; AAB98230.1; -
DR      TIGR; MJO242; -
KW      Ribosomal protein; Complete proteome.
SQ      SEQUENCE 22 AA: 3089 MW; F948D74DD4DC98CB CRC64;

Query Match          20.1%; Score 29; DB 1; Length 22;
Best Local Similarity 41.7%; Pred. No. 4.3e+02;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY      3 RRRRRORRRSGS 14
DB      11 RRLRRRRARS 22
      :|:|:|:|
      :|:|:|:|

RESULT 12
TL16_SPIOL STANDARD: PRT; 29 AA.
AC      P81834;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Thylakoid lumenal 16.5 kDa protein (P16.5) (Fragment).
OS      Spinacia oleracea (Spinach).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
OC      Caryophyllales; Caryophyllales; Chenopodiaceae; Spinacia.
OX      NCBI_TaxID=5562;
RN      [1]
RP      SEQUENCE.
RC      TISSUE=Leaf;
RX      MEDLINE=98175931; Pubmed=9506969;
RA      Kieselbach T., Hagman A., Andersson B., Schroeder W.P.;
RT      "The thylakoid lumen of chloroplasts - Isolation and
RT      characterization."
RL      J. Biol. Chem. 273:6710-6716(1998).
CC      -1- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
KW      Chloroplast; Thylakoid.
FT      NON_TER 29
SQ      SEQUENCE 29 AA: 3464 MW; 58B785764E2623E3 CRC64;

Query Match          19.4%; Score 28; DB 1; Length 29;
Best Local Similarity 55.6%; Pred. No. 8.2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY      2 ERKRRORR 10
DB      18 DRKKRLRQ 26
      :|:|:|:|
      :|:|:|:|

RESULT 13
NEUT_CAVPO STANDARD: PRT; 13 AA.
AC      P32560;
DT      01-OCT-1993 (Rel. 27, Created)
DT      01-OCT-1993 (Rel. 27, Last sequence update)
DT      01-MAR-2002 (Rel. 41, Last annotation update)
DE      Neurolestin (NT).
OS      Cavia porcellus (Guinea pig).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Hystriocognathi; Caviidae; Cavia.
GN      NTS.
RA

```

```

OX NCBTaxID=10141;
RN [1]
RP SEQUENCE.
RC TISSUE=Small intestine; PubMed:3087775;
RX MEDLINE=66248085; PubMed:3087775;
RA Shaw C., Thim L., Conlon J.M.;
RT [Ser]Ineurotensin: isolation from guinea pig intestine.*;
RL FEBS Lett. 202;187-192(1986).
CC -1- FUNCTION: Smooth muscle-contracting peptide.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE NEUROTENSIN FAMILY.
DR PIR: A53608; A53608.
KW Vasoactive.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
SQ SEQUENCE 13 AA; 1680 MW; 4C8314644C4115B3 CRC64;

Query Match 18.8%; Score 27; DB 1; Length 13;
Best Local Similarity 71.4%; Pred. No. 4.8e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YERKKR 7
DB 3 YENKSR 9

RESULT 14
UC06_MAIZE STANDARD; PRT; 15 AA.
ID UC06_MAIZE
AC P80612;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Unknown protein from 2D-page of etiolated coleoptile (Spot 1131)
DE (fragment).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBTaxID=4577;
RN [1]
RP SEQUENCE.
RC TISSUE=Coleoptile;
RA Touzet P., Riccardi F., Morin C., Damerval C., Huet J.-C.,
RA Perrotlet J.-C., Zivy M., de Vienne D.;
RT *The maize two dimensional gel protein database: towards an integrated
RT genome analysis program.*;
RL Theor. Appl. Genet. 93:997-1005(1996).
CC -1- MISCELLANEOUS: ON THE 2D-GEL, THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 6.8, ITS MW IS: 71.0 kDa.
DR Maize-2DPAGE; P60612; COLEOPTILE.
DR MaizeDB: 123930; -.
FT NON_TER 1 1
FT NON_TER 15 15
SQ SEQUENCE 15 AA; 1390 MW; 7005E22830F23D61 CRC64;

Query Match 18.8%; Score 27; DB 1; Length 15;
Best Local Similarity 66.7%; Pred. No. 5.6e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 10 RRGSGTDA 18
DB 7 RRGSGCPDA 15

RESULT 15
HM12_CAEL STANDARD; PRT; 27 AA.
ID HM12_CAEL
AC P17487;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)

```

```

DE Homeobox protein ceh-12 (fragment).
GN Ceh-12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBTaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90245646; PubMed=1970877;
RA Schaller D., Wittmann C., Spicher A., Mueller F., Tobler H.;
RT "Cloning and analysis of three new homeobox genes from the nematode
RT Caenorhabditis elegans.";
RL Nucleic Acids Res. 18:2033-2036(1990).
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17076; CA34928.1; -.
DR PIR: S09504; S09504.
DR HSSP: P02836; IENH.
DR TRANSFAC: T02987; -.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR PROSITE: PS00027; HOMEBOX_1; PARTIAL.
DR PROSITE: PSS0071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT NON_TER 1 1
FT DNA_BIND 1 1 HOMEBOX.
FT NON_TER 27 27
SQ SEQUENCE 27 AA; 3182 MW; FDAFD5D16F3870A5 CRC64;

Query Match 18.8%; Score 27; DB 1; Length 27;
Best Local Similarity 27.3%; Pred. No. 1e+03;
Matches 6; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

OY 3 RKRQRRRSGTDAVAPS 24
DB 6 QNRKRNKRCPSSTPIQSTSS 27

```

Search completed: July 15, 2002, 13:57:02  
 Job time: 315 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 15, 2002, 13:57:48 ; Search time 39.48 Seconds  
(without alignments)  
127.073 Million cell updates/sec

Title: US-09-712-819A-10  
144  
Sequence: 1 YERKKRRQRRRSGTDPATAPASSLOAED 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 12865

Minimum DB seq length: 0  
Maximum DB seq length: 29

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_19:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.rodent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	30.6	26	13	090Y47
2	44	30.6	26	13	090V43
3	44	30.6	26	13	090V47
4	42	29.2	23	8	035223
5	47	25.7	13	12	065331
6	35	24.3	25	13	090V57
7	33	22.9	27	13	057542
8	33	22.9	27	13	057543
9	33	22.9	27	13	057543
10	32	22.2	29	6	09TSS7
11	31	21.5	22	12	085617
12	31	21.5	22	12	085664
13	31	21.5	23	5	085496
14	31	21.5	23	5	017073
15	30	20.8	23	5	094781
16	30	20.8	23	11	035829

17	30	20.8	26	4	09UGK1
18	30	20.8	27	13	09ORS7
19	29.5	20.5	22	13	0919P3
20	29	20.1	26	4	09UNE8
21	29	20.1	26	12	056497
22	29	20.1	28	13	091395
23	29	20.1	29	15	091Q15
24	28	19.4	15	13	09PRM3
25	28	19.4	21	12	09UG30
26	28	19.4	21	12	09UG26
27	28	19.4	21	12	09UG23
28	28	19.4	23	8	09T256
29	28	19.4	23	8	095770
30	28	19.4	23	8	095758
31	28	19.4	26	2	033971
32	28	19.4	26	11	099M13
33	28	19.4	27	2	09FBO1
34	28	19.4	27	5	095X71
35	28	19.4	28	2	09S183
36	28	19.4	28	13	09PRQ3
37	28	19.4	29	15	P89816
38	28	19.4	29	15	P89821
39	27.5	19.1	27	11	09R1H4
40	27	18.8	17	6	09TRY8
41	27	18.8	17	11	064450
42	27	18.8	22	8	0957T4
43	27	18.8	22	13	013047
44	27	18.8	23	12	065291
45	27	18.8	23	4	096PE8

## ALIGNMENTS

RESULT 1					
ID 090Y47	PRELIMINARY;	PRT;	26 AA.		
AC 090Y47;					
DT 01-DEC-2001 (TREMBLrel. 19, Created)					
DR 01-DEC-2001 (TREMBLrel. 19, Last sequence update)					
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)					
DE CHROMO-HELICASE DNA BINDING PROTEIN-1W (FRAGMENT).					
GN CHD1-W.					
OS Cynopsitta spixii x Ara macroura.					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Archosauria; Aves; Neognathae; Psittaciformes; Psittacidae.					
OX NCBI_TaxID=157078;					
RN [1]					
RP SEQUENCE FROM N.A.					
RA Miyaki C.Y., Faria P.J., Griffiths R., Araujo J.C.C., Barros Y.M.;					
RT "The last Spix's Macaw and an Illiger's Macaw produced a hybrid.";					
RL Conserv. Genet. 2:53-55(2001).					
DR EMBL; AF276769; AAK96044.1; -.					
KW Helicase.					
FT NON_TER	1	1			
FT NON_TER	26	26			
SQ SEQUENCE	26 AA;	2973 MW;	F179E3EB84AC1297	CRG64;	
Query Match	30.6%;	Score 44;	DB 13;	Length 26;	
Best local Similarity	52.9%;	Pred. No. 18;			
Matches	9;	Conservative	3;	Mismatches	5;
				Indels	0;
				Gaps	0;
QY 2 ERKKRRRRSGSGTDA 18					
DB 7 ERKRRSRRYSGSDSDS 23					
RESULT 2					
ID 090V43	PRELIMINARY;	PRT;	26 AA.		
AC 090V43;					
DT 01-DEC-2001 (TREMBLrel. 19, Created)					

DT 01-DEC-2001 (TREMblrel. 19, last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
 DE CHROMO-HELICASE DNA BINDING PROTEIN-1W (FRAGMENT).  
 GN CHD1-W.  
 OS Ara macacana.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Psittaciformes; Psittacidae; Ara.  
 OX NCBI\_TaxID=157077;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1, AND 2;  
 RA Miyaki C.Y., Faria P.J., Griffiths R., Araujo J.C.C., Barros Y.M.;  
 RT "The last Spix's Macaw and an Illiger's Macaw produced a hybrid.";  
 RL Conserv. Genet. 2:53-55(2001).  
 DR EMBL: AF276764; AAK96039.1; -;  
 DR EMBL: AF276766; AAK96041.1; -;  
 KW Helicase.  
 FT NON\_TER  
 FT SEQUENCE 26 AA; 2973 MW; F179E3EB84AC1297 CRC64;

Query Match 30.6%; Score 44; DB 13; Length 26;  
 Best Local Similarity 52.9%; Pred. No. 18;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 ERKRRRRRRSGSGTDA 18  
 DB 7 ERHRSRRRYSGSDSDS 23

RESULT 3  
 ID 090VV2 PRELIMINARY; PRT; 26 AA.  
 AC 090VV2;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE CHROMO-HELICASE DNA BINDING PROTEIN-1W (FRAGMENT).  
 GN CHD1-W.  
 OS Cyanopsitta spixii.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Psittaciformes; Psittacidae;  
 OC Cyanopsitta.  
 OX NCBI\_TaxID=51902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1, AND 2;  
 RA Miyaki C.Y., Faria P.J., Griffiths R., Araujo J.C.C., Barros Y.M.;  
 RT "The last Spix's Macaw and an Illiger's Macaw produced a hybrid.";  
 RL Conserv. Genet. 2:53-55(2001).  
 DR EMBL: AF276767; AAK96042.1; -;  
 DR EMBL: AF276768; AAK96043.1; -;  
 KW Helicase.  
 FT NON\_TER  
 FT SEQUENCE 26 AA; 2973 MW; F179E3EB84AC1297 CRC64;

Query Match 30.6%; Score 44; DB 13; Length 26;  
 Best Local Similarity 52.9%; Pred. No. 18;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 ERKRRRRRRSGSGTDA 18  
 DB 7 ERHRSRRRYSGSDSDS 23

RESULT 4  
 ID 035223 PRELIMINARY; PRT; 23 AA.  
 AC 035223;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
 DE RIBOSOMAL PROTEIN S19 (FRAGMENT).  
 GN RPS19.  
 OS Oenothera lutea (Bertero's evening primrose).  
 OC Mitochondrion.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Myricales; Onagraceae; Oenothera.  
 OX NCBI\_TaxID=3930;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94250844; PubMed=8193306;  
 RA Beck H., Brennicke A., Schuster W.;  
 RT "rps3 and rpl16 genes do not overlap in Oenothera mitochondria: GNG  
 as a potential translation initiation codon in plant mitochondria?";  
 RL Plant Mol. Biol. 24:811-818(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92107682;  
 RA Schuster W., Brennicke A.;  
 RT "RNA editing makes mistakes in plant mitochondria: editing looses  
 RT sense in transcripts of a rps19 pseudogene and in creating stop codons  
 RT in coxi and rpl16 mRNAs of Oenothera.";  
 RL Nucleic Acids Res. 19:6923-6928(1991).  
 DR EMBL: X69140; CAA48893.1; -;  
 KW Mitochondrion.  
 FT NON\_TER  
 FT SEQUENCE 23 AA; 2764 MW; 08B5AB73D448C5CF CRC64;

Query Match 29.2%; Score 42; DB 8; Length 23;  
 Best Local Similarity 53.3%; Pred. No. 30;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 YERKRRRRRRSGSG 15  
 DB 3 FTRKRRRRRTNIGSG 17

RESULT 5  
 ID 065331 PRELIMINARY; PRT; 13 AA.  
 AC 065331;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-NOV-1996 (TREMblrel. 08, Last annotation update)  
 DE BASIC PROTEIN (FRAGMENT).  
 OS Autographa californica nuclear polyhedrosis virus (AcMNPV).  
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
 OC Nucleopolyhedrovirus.  
 OX NCBI\_TaxID=46015;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C6;  
 RX MEDLINE=90218045; PubMed=2109042;  
 RA Hill-Perkins M.S., Possee R.D.;  
 RT "A baculovirus expression vector derived from the basic protein  
 RT promoter of Autographa californica nuclear polyhedrosis virus.";  
 RL J. Gen. Virol. 71:971-976(1990).  
 DR EMBL: D00864; BAA00738.1; -;  
 FT NON\_TER  
 FT SEQUENCE 13 AA; 1626 MW; 6C75B58A749D1414 CRC64;

Query Match 25.7%; Score 37; DB 12; Length 13;  
 Best Local Similarity 70.0%; Pred. No. 85;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 7 RQRRRRSGGT 16  
 DB 4 RRRRRSGGT 13



RESULT 6  
 Q90Y57 PRELIMINARY; PRT; 25 AA.  
 AC Q90Y57;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE RIBOSOMAL PROTEIN L41.  
 OS Ictalurus punctatus (Channel catfish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Siluriformes; Ictaluridae; Ictalurus.  
 NCBI\_TaxID=7998;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Patterson A.P., Karsi A., Liu Z.J.;  
 RT "Translational machinery of channel catfish: II. Complementary DNA and  
 RT Expression of the complete set of 47 60S Ribosomal Proteins.";  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF401598; AAK95170.1;  
 SQ SEQUENCE 25 AA; 3357 MW; 9AD629C69EDC029E CRC64;

Query Match 24.3%; Score 35; DB 13; Length 25;  
 Best Local Similarity 54.5%; Pred. No. 3.2e+02;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 ERKRRRRRRS 12  
 :|||:|:|  
 Db 14 KRRRRRRRRS 24

RESULT 7  
 Q962S2 PRELIMINARY; PRT; 25 AA.  
 AC Q962S2;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE RIBOSOMAL PROTEIN L41.  
 OS Spodoptera frugiperda (Fall armyworm).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Noctuidae; Noctuidae; Amphipyrinae; Spodoptera.  
 NCBI\_TaxID=7108;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Landais I., Ogilastro M., Mita K., Nohata J., Lopez-Ferber M.,  
 RA Duport-Cerutti M., Fournier P., Devauchelle G.;  
 RT "Full-length ribosomal protein sequence from an EST library of  
 RT Spodoptera frugiperda cells (Sf9)."  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF400204; AAK92176.1;  
 SQ SEQUENCE 25 AA; 3399 MW; 9AD629D9ECC039E CRC64;

Query Match 22.9%; Score 33; DB 5; Length 25;  
 Best Local Similarity 54.5%; Pred. No. 6.1e+02;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 ERKRRRRRRS 12  
 :|||:|:|  
 Db 14 KRRRRRRRRS 24

RESULT 8  
 Q57542 PRELIMINARY; PRT; 27 AA.  
 AC Q57542;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE HOMEBOX PROTEIN LPHOX1B (FRAGMENT).  
 OS Lampetra planeri (Brook lamprey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;  
 OC Petromyzontiformes; Petromyzontidae; Lampetra.  
 NCBI\_TaxID=7750;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98358009; PubMed=9694633;  
 RA Sherman A.C., Holland P.W.;  
 RT "Estimation of Hox gene cluster number in lampreys.";  
 RL Int. J. Dev. Biol. 42:617-620(1998).  
 DR EMBL; AF044798; AAC03002.1;  
 FT NON\_TER 1  
 FT NON\_TER 27  
 SQ SEQUENCE 27 AA; 3098 MW; FE1EE5E971D0C1DA CRC64;

Query Match 22.9%; Score 33; DB 13; Length 27;  
 Best Local Similarity 46.7%; Pred. No. 6.6e+02;  
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 RKRRRRRSGSGTD 17  
 :|||:|:|  
 Db 9 RAPRRRRRAAARD 23

RESULT 9  
 Q57543 PRELIMINARY; PRT; 27 AA.  
 AC Q57543;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HOMEBOX PROTEIN LPHOX1C (FRAGMENT).  
 OS Lampetra planeri (Brook lamprey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;  
 OC Petromyzontiformes; Petromyzontidae; Lampetra.  
 NCBI\_TaxID=7750;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98358009; PubMed=9694633;  
 RA Sherman A.C., Holland P.W.;  
 RT "Estimation of Hox gene cluster number in lampreys.";  
 RL Int. J. Dev. Biol. 42:617-620(1998).  
 DR EMBL; AF044799; AAC03003.1;  
 FT NON\_TER 1  
 FT NON\_TER 27  
 SQ SEQUENCE 27 AA; 3190 MW; B86EE73971D0DC06 CRC64;

Query Match 22.9%; Score 33; DB 13; Length 27;  
 Best Local Similarity 46.7%; Pred. No. 6.6e+02;  
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 RKRRRRRSGSGTD 17  
 :|||:|:|  
 Db 9 RATRRRRRAATERD 23

RESULT 10  
 Q9TSS7 PRELIMINARY; PRT; 29 AA.  
 AC Q9TSS7;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE WILM'S TUMOR PROTEIN 1 (FRAGMENT).  
 GN WT1.  
 OS Felis silvestris catus (Cat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
 NCBI\_TaxID=9685;  
 RN [1]

```

RP SEQUENCE FROM N.A.
RX MEDLINE-21015404; PubMed-11130975;
RA Brouillette J.A., Andrew J.R., Venta P.J.;
RT "Estimate of nucleotide diversity in dogs with a pool-and-sequence
method."
RL Mamm. Genome 11:1079-1086(2000).
DR EMBL; AF201740; AAF19825.1; -.
DR EMBL; AF201739; AAF19825.1; JOINED.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 29 AA; 3594 MW; FDC8E57A0BAE35DD CRC64;

Query Match 22.2%; Score 32; DB 6; Length 29;
Best Local Similarity 66.7%; Pred. No. 9.9e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 5 KRRORRRSG 13
DB 1 KRRORRRHG 9

RESULT 11
O85617 PRELIMINARY; PRT; 20 AA.
AC 085617;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE REOVIRUS SEROTYPE 1 L3 (FRAGMENT).
OC Reovirus (type 1 / strain Lang).
OS Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.
OX NCBI_TaxID=10884;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-83017877; PubMed-7123853;
RA Galliard R.K., Li J.K., Keene J.D., Joklik W.K.;
RT "The sequences at the termini of four genes of the three reovirus
serotypes."
RL Virology 121:320-326(1982).
DR EMBL; J02297; AAA47233.1; -.
FT NON_TER 20
FT NON_TER 20
SQ SEQUENCE 20 AA; 2177 MW; FA861D930E1FD8BA CRC64;

Query Match 21.5%; Score 31; DB 12; Length 20;
Best Local Similarity 33.3%; Pred. No. 9.4e+02;
Matches 5; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 5 KRRORRRSGSGTDTAT 19
DB 6 RKTGKSSGKNDST 20

RESULT 12
O85664 PRELIMINARY; PRT; 22 AA.
AC 085664;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE MAJOR CORE PROTEIN LAMBDA-1 (FRAGMENT).
OC Reovirus (type 3 / strain Deering).
OS Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.
OX NCBI_TaxID=10886;
RN [1]
RP SEQUENCE OF 1-6 FROM N.A.
RX STRAIN-DEARING;
RX MEDLINE-82217029; PubMed-7086967;
RA Kozak M.;
RT "Sequences of ribosome binding sites from the large size class of
J. Virol. 42:467-473(1982).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-DEARING;
RX MEDLINE-83017876; PubMed-6927854;
RA Antczak J.B., Chmelo R., Pickup D.J., Joklik W.K.;
RT "Sequences at both termini of the 10 genes of reovirus serotype 3
(R strain Deering)."
RL Virology 121:307-319(1982).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-DEARING;
RX MEDLINE-83017877; PubMed-7123853;
RA Galliard R.K., Li J.K., Keene J.D., Joklik W.K.;
RT "The sequences at the termini of four genes of the three reovirus
serotypes."
RL Virology 121:320-326(1982).
DR EMBL; J02317; AAA47254.1; -.
FT NON_TER 22
FT NON_TER 22
SQ SEQUENCE 22 AA; 2462 MW; D6706A861D930E1F CRC64;

Query Match 21.5%; Score 31; DB 12; Length 22;
Best Local Similarity 33.3%; Pred. No. 1e+03;
Matches 5; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 5 KRRORRRSGSGTDTAT 19
DB 6 RKTGKSSGKNDST 20

RESULT 13
O85496 PRELIMINARY; PRT; 23 AA.
AC 085496;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE ENVELOPE PROTEIN (FRAGMENT).
OC Bovine leukemia virus (BLV).
OS Viruses; Retroid viruses; Retroviridae; Deltaretrovirus.
OX NCBI_TaxID=11901;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-87198865; PubMed-3033284;
RA Rice N.R., Simek S.L., Dubois G.C., Showalter S.D., Gilden R.V.,
RA Stephens R.M.;
RT "Expression of the bovine leukemia virus X region in virus-infected
cells."
RL J. Virol. 61:1577-1585(1987).
DR EMBL; M16017; AAA87336.1; -.
FT NON_TER 23
FT NON_TER 23
SQ SEQUENCE 23 AA; 2877 MW; F85E7CEBD5440B08 CRC64;

Query Match 21.5%; Score 31; DB 15; Length 23;
Best Local Similarity 75.0%; Pred. No. 1.1e+03;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 KRRORRR 11
DB 3 KERRRRR 10

RESULT 14
O17073 PRELIMINARY; PRT; 29 AA.
AC 017073;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE SLKMOH (A. POLYPHEMUS) CHORION PROTEIN (FRAGMENT).
OS Anthraea polyphemus (Polypheumus moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

```

OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Bombycoidea; Saturniidae; Saturniinae; Anthrea.
OX NCBI_TaxID=7120;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80090072; Pubmed=519771;
RA Jones C.W., Rosenthal N., Rodakis G.C., Kafatos F.C.;
RT "Evolution of two major chorion multigene families as inferred from
  cloned cDNA and protein sequences."
RL Cell 18:1317-1332(1979).
DR EMBL; K01842; AAA27785.1; .
KW Chorion.
FT NON_TER 1 1
FT NON_TER 29 29
SQ SEQUENCE 29 AA; 2655 MW; 26483F783526E820 CRC64;

Query Match 21.5%; Score 31; DB 5; Length 29;
Best Local Similarity 41.2%; Pred. No. 1.4e+03;
Matches 7; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY 11 RSGSGTDATAPASSIOA 27
   | | | : | |
Db 5 RLCCCGGCGISPAEILAA 21

RESULT 15
O94781 PRELIMINARY; PRT; 23 AA.
AC O94781:
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
DE HISTONE H2A (FRAGMENT).
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Tanaka T.; Tanaka M.;
RT "The parasite genome project: Trypanosoma cruzi."
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; D87227; BAA13318.1; -.
FT NON_TER 1 1
FT NON_TER 1 1
SQ SEQUENCE 23 AA; 2790 MW; 12E9ED7592E52045 CRC64;

Query Match 20.8%; Score 30; DB 5; Length 23;
Best Local Similarity 66.7%; Pred. No. 1.5e+03;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 RKKRQRRR 11
   | | | : |
Db 10 RDKRRRRR 16

```

Search completed: July 15, 2002, 13:57:49  
 Job time: 337 sec

10

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:51:41 ; Search time 48.92 Seconds

(Without alignments)  
65.845 Million cell updates/sec

Title: US-09-712-819A-10

Perfect score: 144  
Sequence: I YEKRRRRRRSGSGTDAFAPASSLQABD 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 272466

Minimum DB seq length: 0  
Maximum DB seq length: 29

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	39.6	25	AA867911	HIV-1 tat derived
2	56	38.9	18	AA866601	HIV TAT cellular u
3	55	38.2	15	AA89577	Residues tat47-586
4	55	38.2	15	AAW31207	HIV-1 based transp
5	55	38.2	15	AAW32916	HIV-1 tat protein
6	55	38.2	15	AAW26448	HIV-1 tat protein fr
7	55	38.2	15	AAW76153	HIV Type I TAT pro
8	55	38.2	16	AAW19818	Universal transfer
9	55	38.2	22	AAE12486	Membrane transiti
10	55	38.2	22	AAE12487	Membrane transiti
11	55	38.2	24	AAE12487	Residues 37-586GC

12	55	38.2	24	18	AAW31205	HIV-1 based transp
13	55	38.2	24	18	AAW32914	HIV-1 tat protein
14	55	38.2	24	18	AAW26446	HIV tat protein
15	55	38.2	24	19	AAW76151	HIV Type I TAT pro
16	55	38.2	26	15	AAW49558	Residues 47-72 HIV
17	55	38.2	26	18	AAW31211	HIV-TAT protein tr
18	55	38.2	26	18	AAW26451	Peptide of the inv
19	55	38.2	26	19	AAW76157	HIV Type I TAT pro
20	55	38.2	27	20	AAW00349	Fragment of human
21	54	37.5	25	15	AAW49567	Residues 38-62 HIV
22	54	37.5	25	18	AAW31216	HIV-TAT protein tr
23	54	37.5	25	18	AAW26456	Peptide of the inv
24	54	37.5	25	19	AAW76162	HIV Type I TAT pro
25	54	37.5	26	11	AAW03434	AAW37-62 of HIV t
26	54	37.5	26	11	AAW02238	Tat-dm37-62, 46A1a
27	54	37.5	26	11	AAW02239	Tat-dm37-62, 46A1a
28	54	37.5	26	11	AAW04720	Tat-dm37-62, 40A1a
29	54	37.5	26	11	AAW02240	Tat-dm37-62, 39A1a
30	54	37.5	26	11	AAW02241	Tat-dm37-62, 38A1a
31	54	37.5	26	11	AAW02241	Tat-dm37-62, 37A1a
32	54	37.5	26	11	AAW00086	pH-dependent membr
33	54	37.5	26	21	AAW96570	HIV transactivator
34	54	37.5	29	19	AAW56399	Preferred signal s
35	54	37.5	29	21	AAW55820	Signal sequence fo
36	54	37.5	29	22	AAW11950	Membrane transport
37	53	36.8	13	21	AAW69771	RNA binding peptid
38	53	36.8	13	22	AAW06108	Peptide associated
39	53	36.8	21	22	AAW55701	Tat-CLASP1carboxyl
40	53	36.8	21	22	AAW57548	Tat-CLASP1carboxyl
41	53	36.8	21	22	AAW57958	Human p15NTR depen
42	53	36.8	28	20	AAW33518	Beta-catenin modul
43	52.5	36.5	22	21	AAW03933	HIV Tat protein fr
44	52.5	36.5	24	13	AAW27034	Peptide associated
45	52	36.1	21	22	AAW55700	

#### ALIGNMENTS

RESULT 1

AA867911  
ID AAB67911 standard; peptide: 25 AA.  
XX  
AC AAB67911;  
XX  
DT 29-JUN-2001 (first entry)  
XX  
XX HIV-1 tat derived peptide which is rich in basic amino acids.  
DE  
XX  
XX Heparin affinity regulatory peptide; HARP; angiogenic peptide; HIV;  
KW Immune response; cellular growth; cellular regeneration; muscle cell;  
KW cistricisation; immunosuppressive disease; human immunodeficiency virus;  
KW HIV infection; antiviral.

OS Human immunodeficiency virus type 1.  
XX  
XX WO200127136-A2.  
XX  
XX  
XX 19-APR-2001.  
XX  
XX  
XX 06-OCT-2000; 2000WO-FR02786.  
XX  
XX 12-OCT-1999; 99FR-0012714.  
XX  
XX  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
XX Barilault D, Achour A, Courty J;  
XX  
XX WPI; 2001-281970/29.  
XX  
XX  
XX New Heparin Affinity Regulatory Peptide angiogenic peptides, useful for  
PT regenerating muscle cells, aiding cistricisation, detecting and  
PT treating immunosuppressive diseases, stimulate immune response and

PT tissue regeneration -  
XX  
PS Disclosure; Page 3; 43pp; French.  
XX  
CC AAB67904-16 represent peptides which rich in basic amino acids. The  
CC specification describes heparin affinity regulatory peptide (HARP)  
CC angiogenic peptides that stimulate immune response, cellular growth  
CC and regeneration. The peptides are useful for regenerating cells, such  
CC as muscle cells, and in aiding cicatrization, as well as in the  
CC treatment of immunosuppressive diseases. The peptides amplify the  
CC replication of human immunodeficiency virus (HIV) in vitro, and are  
CC useful in the detection of HIV infections. When used with anti-viral  
CC agents, they render the HIV more accessible to the antiviral agent,  
CC and therefore are more easily destroyed.  
XX  
SQ Sequence 25 AA:  
  
Query Match 39.6%; Score 57; DB 22; Length 25;  
Best Local Similarity 66.7%; Pred. No. 0.4;  
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
QY 1 YERKKRRRRSGGTDA 18  
| | | | | | | | | | | | | | | | | | | | | |  
Db 7 ygrkkrrrrrrppgnga 24  
  
RESULT 2  
ID AAR86601 standard; peptide; 18 AA.  
XX  
AC AAR86601;  
XX  
DT 28-JUN-1996 (first entry)  
XX  
DE HIV Tat cellular uptake region immunogen #1.  
XX  
KW Immunogen: cellular uptake region; transactivating protein; Tat; HIV;  
XX  
OS viraemia; antibody.  
XX  
XX Human immunodeficiency virus type 1.  
XX  
PN MO9531999-A1.  
XX  
PD 30-NOV-1995.  
XX  
PF 16-MAY-1995; 95WO-US06077.  
XX  
PR 23-MAY-1994; 94US-0247991.  
XX  
PA (IMMU-) IMMUNOCIOLOGY RES INST INC.  
XX  
PI Culler MD, Goldstein G, Shenbagamurthi P;  
XX  
DR WPI: 1996-049298/05.  
XX  
PT Immunogen containing cellular uptake region of viral Tat protein -  
PT induces high antibody titre against Tat, partic. for preventing or  
PT treating HIV infection  
XX  
PS Claim 6; Page 45; 54pp; English.  
XX  
CC The sequences given in AAR86601-19 are immunogens derived from the  
CC cellular uptake region of the transactivating protein, Tat of HIV.  
CC These peptides are based on the region comprising amino acids 46-63  
CC (see AAR86625) of the HIV Tat protein. These peptides are used to  
CC protect against infection by HIV. They may also be used to reduce  
CC viraemia in already infected patients. These immunogens may be used  
CC to identify, or generate, antibodies specific for Tat.  
XX  
SQ Sequence 18 AA:

Query Match 38.9%; Score 56; DB 17; Length 18;  
Best Local Similarity 68.8%; Pred. No. 0.39; Mismatches 3; Indels 0; Gaps 0;  
Matches 11; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 YERKKRRRRSGGT 16  
| | | | | | | | | | | | | | | | | | | | | |  
Db 2 ygrkkrrrrrrppgs 17  
  
RESULT 3  
ID AAR9577 standard; peptide; 15 AA.  
XX  
AC AAR9577;  
XX  
DT 23-SEP-1994 (first entry)  
XX  
DE Residues tat47-58GCC HIV tat protein.  
XX  
KW HIV; tat; transport; molecy; conjugate; cargo molecule;  
KW cytoplasmic delivery; nuclear delivery; cysteine-rich region;  
KW transactivation; disulphide aggregation.  
XX  
OS Human immunodeficiency virus type 1.  
XX  
PN MO9404686-A.  
XX  
PD 03-MAR-1994.  
XX  
PF 19-AUG-1993; 93WO-US07833.  
XX  
PR 21-AUG-1992; 92US-0934375.  
XX  
PA (BIOJ ) BIOGEN INC.  
XX  
PI Barsom JG, Fawell SE, pepinsky RB;  
XX  
DR WPI: 1994-083202/10.  
XX  
PT Transport of cargo moieties into cells - using genetic fusions of  
PT chemical conjugates comprising a portion of HIV tat protein as  
PT transport peptide  
XX  
PS Disclosure; Page 87; 153pp; English.  
XX  
CC The sequences given in AAR9574-77 are fragments of the HIV tat protein  
CC which were used as transport moieties in the conjugate of the  
CC invention. These conjugates are used as cargo molecules for the  
CC efficient cytoplasmic and nuclear delivery of biologically active  
CC non-tat proteins, nucleic acids and other molecules which are not  
CC inherently capable of entering a target cell. The tat basic region  
CC amino acid sequence carries cargo molecules by covalent attachment.  
CC The reduced size of the transport peptides minimises interference  
CC with the biological activity of the cargo molecule. By virtue of  
CC the absence of the cysteine-rich region (residues 22-36) of the tat  
CC protein, problems of spurious transactivation and disulphide  
CC aggregation are solved. The reduced transport peptide size also  
CC enhances uptake efficiency.  
XX  
SQ Sequence 15 AA:  
  
Query Match 38.2%; Score 55; DB 15; Length 15;  
Best Local Similarity 84.6%; Pred. No. 0.45; Mismatches 2; Indels 0; Gaps 0;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 YERKKRRRRSG 13  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 ygrkkrrrrrrp 13  
  
RESULT 4  
AAW31207

ID AAW31207 standard; peptide; 15 AA.  
 XX AAW31207;  
 AC  
 DT 20-MAR-1998 (first entry)  
 XX  
 DE HIV-1 based transport polypeptide Tat47-58GSC.  
 XX  
 KW Human immunodeficiency virus; HIV Type 1; Tat protein;  
 KW cargo molecules; intracellular delivery; fusion protein;  
 KW therapeutic; prophylactic; diagnostic; transport polypeptide.  
 XX  
 OS Synthetic.  
 OS Human immunodeficiency virus type 1.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 3..11  
 FT /label= basic\_region  
 FT Misc-difference 13..15  
 FT /label= spacer\_residues  
 FT  
 XX US5674980-A.  
 XX  
 PD 07-OCT-1997.  
 XX  
 PD 21-DEC-1989; 89US-0454450.  
 XX  
 PF 28-APR-1994; 94US-0235403.  
 PR 21-DEC-1989; 89US-0454450.  
 PR 02-JAN-1991; 91US-0636662.  
 PR 21-AUG-1992; 92US-0934375.  
 PR 19-AUG-1992; 92US-0934375.  
 PR 19-AUG-1992; 93US-0507833.  
 PR 24-NOV-1993; 93US-0158015.  
 PR 25-MAY-1995; 95US-0450098.  
 XX  
 PA (BARS/) BARSOUM J G.  
 PA (FAWE/) FAWELL S E.  
 PA (FRAN/) FRANKEL A.  
 PA (PABO/) PABO.  
 PA (PEPI/) PEPINSKY R B.  
 XX  
 PI Barsoum JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;  
 XX  
 DR WPI; 1997-502388/46.  
 XX  
 PT Fusion proteins containing truncated HIV tat sequences - useful for  
 PT intracellular delivery of viral repressor proteins  
 XX  
 PS Example 1; Column 55; 77pp; English.  
 XX  
 CC This sequence represents the transport polypeptide Tat47-58GSC which can  
 CC be used in a novel method for the delivery of biologically active cargo  
 CC molecules into the cytoplasm and nuclei of eukaryotic cells. The tat  
 CC protein from immunodeficiency virus (e.g. HIV-1, HIV-2, SIV) is readily  
 CC taken up into cells when present extracellularly. This polypeptide  
 CC has been constructed from the HIV-1 tat protein (see AAW31202) by  
 CC removing the cysteine-rich region and the carboxy-terminal domain while  
 CC retaining the basic region. Cargo molecules could be covalently linked  
 CC to the protein to form a fusion protein without the existing problems of  
 CC spurious trans-activation and disulphide aggregation. The reduced size  
 CC of transport polypeptides also minimises interference with the  
 CC biological activity of the cargo molecule. This is applicable for  
 CC therapeutic, prophylactic or diagnostic intracellular delivery of small  
 CC molecules and macromolecules e.g. proteins, nucleic acids and  
 CC polysaccharides.  
 CC  
 XX Sequence 15 AA;  
 SQ

Query Match 38.2%; Score 55; DB 18; Length 15;  
 Best Local Similarity 84.6%; Pred. No. 0.45;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YEKRRRRRRSG 13  
 1 | | | | | | | | | |  
 Db 1 ygkkrqrrrrp 13  
 RESULT 5  
 AAW32916  
 ID AAW32916 standard; peptide; 15 AA.  
 XX  
 AC AAW32916;  
 XX  
 DT 19-JAN-1998 (first entry)  
 XX  
 DE HIV-1 tat protein residues 47-58 plus Gly-Gly-Cys linker.  
 XX  
 KW Chemical conjugate; transport polypeptide; tat protein;  
 KW nucleic acid; delivery; cytoplasm; cell nucleus; therapy;  
 KW prophylaxis; diagnosis; spurious transactivation; HIV-1;  
 KW disulphide aggregation.  
 XX  
 OS Human immunodeficiency virus type 1.  
 XX  
 PN US5670617-A.  
 XX  
 PD 23-SEP-1997.  
 XX  
 PD 21-DEC-1989; 89US-0454450.  
 XX  
 PF 28-APR-1994; 94US-0235403.  
 PR 21-DEC-1989; 89US-0454450.  
 PR 02-JAN-1991; 91US-0636662.  
 PR 21-AUG-1992; 92US-0934375.  
 PR 19-AUG-1992; 93US-0507833.  
 PR 24-NOV-1993; 93US-0158015.  
 PR 25-MAY-1995; 95US-0450246.  
 XX  
 PA (BARS/) BARSOUM J G.  
 PA (FAWE/) FAWELL S E.  
 PA (FRAN/) FRANKEL A.  
 PA (PABO/) PABO C.  
 PA (PEPI/) PEPINSKY R B.  
 XX  
 PI Barsoum JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;  
 XX  
 DR WPI; 1997-479523/44.  
 XX  
 PT Conjugate for intracellular delivery - comprising transport moiety  
 PT having amino acids 49-57 of human immunodeficiency virus tat protein  
 PT and nucleic acid cargo moiety  
 XX  
 PS Disclosure; Columns 57-58; 77pp; English.  
 XX  
 CC A novel covalently linked chemical conjugate, comprises a transport  
 CC polypeptide moiety consisting of at least residues 49-57 of human  
 CC immunodeficiency virus (HIV) tat protein, but not residues 22-36  
 CC and 73-86; e.g. the present peptide, and a cargo moiety comprising  
 CC a single or double stranded nucleic acid, a cargo moiety could be used  
 CC to deliver cargo molecules into the cytoplasm and nuclei of cells  
 CC for therapeutic, prophylactic and diagnostic applications. The HIV  
 CC tat protein is readily taken up into cells and the cell nucleus.  
 CC The reduced size of the transport polypeptides minimises  
 CC interference with the biological activity of the cargo molecule.  
 CC In addition, by virtue of the absence of the cysteine rich region  
 CC of the HIV tat protein, the transport polypeptides solve the  
 CC potential problems of spurious transactivation and disulphide  
 CC aggregation.  
 CC  
 XX Sequence 15 AA;  
 SQ

Query Match 38.2%; Score 55; DB 18; Length 15;  
 Best Local Similarity 84.6%; Pred. No. 0.45;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YERKKRRRRSG 13  
| | | | | | | | | |  
Db 1 Ygrrkrrrrrrpg 13

RESULT 6  
AAW26448  
ID AAW26448 standard; Protein; 15 AA.  
XX  
AC AAW26448;  
XX  
DT 16-DEC-1997 (first entry)  
XX  
DE HIV tat protein fragment (aa47-58GCC).  
XX  
KW HIV; tat protein; transport protein; cargo delivery.  
XX  
OS Chimeric human immunodeficiency virus type 1;  
XX  
OS Chimeric synthetic.  
XX  
PN US5652122-A.  
XX  
PD 29-JUL-1997.  
XX  
PF 21-DEC-1989; 89US-0454450.  
XX  
PR 28-APR-1994; 94US-0235403.  
PR 21-DEC-1989; 89US-0454450.  
PR 02-JAN-1991; 91US-0636662.  
PR 21-AUG-1992; 92US-0934375.  
PR 19-AUG-1993; 93WO-US07833.  
PR 24-NOV-1993; 93US-0158015.  
PR 25-MAY-1995; 95US-0450257.  
XX  
PA (BARS/) BARSOUM J G.  
PA (FAWE/) FAWELL S E.  
PA (FRAN/) FRANKEL A.  
PA (PABO/) PABO C.  
PA (PEPI/) PEPIINSKY R B.  
XX  
PI Barsoum JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;  
XX  
XX WPI: 1997-392943/36.  
XX  
DR  
XX  
PT New DNA constructs for transporting molecules to cells - encode a  
PT fusion protein comprising a modified HIV tat protein and a  
PT carboxy-terminal cargo moiety  
XX  
PS Disclosure: Column 57-58; 76pp; English.  
XX  
CC This polypeptide comprises amino acids 47-58 of the HIV tat  
CC protein (see also AAW26443), plus a Gly-Gly-Cys C-terminal extension.  
CC Novel transport proteins comprise modified HIV tat covalently  
CC attached to a cargo molecule (see AAW26436-42). The modified tat  
CC protein (see AAW26444-45) is characterised by the presence of the tat  
CC basic region (amino acids 47-59), and the absence of the Cys-rich  
CC region (amino acids 22-36) and the exon 2-encoded C-terminal region  
CC (amino acids 73-86). These modifications solve the potential  
CC problems of spurious trans-activation and disulphide aggregation,  
CC while the reduced size of the transport protein minimises  
CC interference with the biological activity of the cargo molecule.  
CC DNA molecules that encode the modified tat fusion proteins are  
CC claimed and can be used to deliver polypeptides or nucleic acids to  
CC the cytoplasm or nuclei of cells in vivo or in vitro.  
XX  
SQ Sequence 15 AA;

Query/Match 38.2%; Score 55; DB 18; Length 15;  
Best Local Similarity 84.6%; Pred. No. 0.45;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YERKKRRRRSG 13  
| | | | | | | | | |  
Db 1 Ygrrkrrrrrrpg 13

RESULT 7  
AAW76153  
ID AAW76153 standard; protein; 15 AA.  
XX  
AC AAW76153;  
XX  
DT 24-NOV-1998 (first entry)  
XX  
DE HIV Type I Tat protein fragment tat47-58GCC.  
XX  
KW Tat protein; cargo molecule; therapy; diagnosis; transport protein;  
KW fusion protein; human papillomavirus E2 repressor; target cell.  
XX  
OS Human immunodeficiency virus type 1.  
XX  
PN US5804604-A.  
XX  
PD 08-SEP-1998.  
XX  
PF 25-MAY-1995; 95US-0450236.  
XX  
PR 28-APR-1994; 94US-0235403.  
PR 21-DEC-1989; 89US-0454450.  
PR 02-JAN-1991; 91US-0636662.  
PR 19-AUG-1993; 93WO-US07833.  
PR 24-NOV-1993; 93US-0158015.  
PR 25-MAY-1995; 95US-0450236.  
XX  
PA (BIOJ ) BIOGEN INC.  
XX  
PI Barsoum JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;  
XX  
XX WPI: 1998-505702/43.  
XX  
DR  
XX  
PT HIV tat-derived transport fusion proteins - used to deliver  
PT biological active molecules e.g. peptide(s) or nucleic acids,  
PT specifically into cytoplasm or nuclei of cells  
XX  
PS Disclosure: Column 57-58; 83pp; English.  
XX  
CC This sequence represents a human immunodeficiency virus (HIV) Type I  
CC Tat protein fragment, tat47-58GCC which is used in a method for the  
CC delivery of biologically active cargo molecules into the cytoplasm and  
CC nuclei of cells, for therapeutic, prophylactic or diagnostic purposes.  
CC This is accomplished by the presence of a small, basic section of tat  
CC transport protein of HIV. This is used as it is this protein which is  
CC observed to cause human cells in culture to take up HIV. The method  
CC involves the use of a cargo moiety in combination with a transport  
CC moiety usually in the form of a fusion protein. The cargo moiety is a  
CC human papillomavirus E2 repressor that retains its biological activity  
CC after delivery into a target cell and where the transport moiety is  
CC one of following HIV tat protein fragments (a) aa 47-58, (b) aa 47-72,  
CC (c) 38-72, (d) aa 38-58, (e) aa 37-58, (f) aa 1-21 and 38-72,  
CC (g) aa 47-62 or aa 38-62. The proteins allow delivery of specific  
CC peptides into cells at high concentrations due to use of existing  
CC transporters. Previous methods of delivery include bombardment and  
CC transforming, which only allow a fraction of the cell population to be  
CC infected and can additionally damage cells as they cause physical opening  
CC of the cell walls/membranes to allow entry.  
XX  
SQ Sequence 15 AA;

Query/Match 38.2%; Score 55; DB 19; Length 15;  
Best Local Similarity 84.6%; Pred. No. 0.45;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;







```

XX (BARS/) BARSOUX J G.
PA (FAME/) FAMELL S E.
PA (FRAN/) FRANKEL A.
PA (PABO/) PABO.
PA (PEPI/) PEPINSKY R B.
XX
PI Barsoum JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;
XX
XX WPI; 1997-502388/46.
XX
XX Fusion proteins containing truncated HIV tat sequences - useful for
XX intracellular delivery of viral repressor proteins
XX
XX Example 1; Column 55; 77pp; English.
XX
PS This sequence represents the transport polypeptide Tat38-58GCC which can
CC be used in a novel method for the delivery of biologically active cargo
CC molecules into the cytoplasm and nuclei of eukaryotic cells. The tat
CC protein from immunodeficiency virus (e.g. HIV-1, HIV-2, SIV) is readily
CC taken up into cells when present extracellularly. This polypeptide
CC has been constructed from the HIV-1 tat protein (see AAMW31202) by
CC removing the cysteine-rich region and the carboxy-terminal domain while
CC retaining the basic region. Cargo molecules could be covalently linked
CC to the protein to form a fusion protein without the existing problems of
CC spurious trans-activation and disulphide aggregation. The reduced size
CC of transport polypeptides also minimises interference with the
CC biological activity of the cargo molecule. This is applicable for
CC therapeutic, prophylactic or diagnostic intracellular delivery of small
CC molecules and macromolecules e.g. proteins, nucleic acids and
CC polysaccharides.
XX
XX Sequence 24 AA:
SQ

```

Query Match 38.2%; Score 55; DB 18; Length 24;  
 Best Local Similarity 84.6%; Pred. No. 0.73;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 YEKRRRRRRRSG 13
   1 | | | | | | | | |
Db 10 YGRKKRRGRRTPG 22

```

RESULT 13  
 AAM32914  
 ID AAM32914 standard; peptide; 24 AA.  
 AC AAM32914;  
 AC  
 DT 19-JAN-1998 (first entry)  
 DT  
 XX HIV-1 tat protein residues 38-58 plus Gly-Gly-Cys linker.  
 DE  
 XX  
 KW Chemical conjugate; transport polypeptide; tat protein;  
 KW nucleic acid; delivery; cytoplasm; cell nucleus; therapy;  
 KW prophylaxis; diagnosis; spurious transactivation; HIV-1;  
 KW disulphide aggregation.  
 XX  
 OS Human immunodeficiency virus type 1.  
 OS  
 PN US5670617-A.  
 PN  
 PD 23-SEP-1997.  
 PD  
 XX 21-DEC-1989; 89US-0454450.  
 XX  
 XX 28-APR-1994; 94US-0235403.  
 PR 21-DEC-1989; 89US-0454450.  
 PR 02-JAN-1991; 91US-0636662.  
 PR 21-AUG-1992; 92US-0934375.  
 PR 19-AUG-1993; 93WO-US07833.  
 PR 24-NOV-1993; 93US-0158015.

```

PR 25-MAY-1995; 95US-0450246.
XX
XX (BARS/) BARSOUX J G.
PA (FAME/) FAMELL S E.
PA (FRAN/) FRANKEL A.
PA (PABO/) PABO C.
PA (PEPI/) PEPINSKY R B.
XX
PI Barsoum JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;
XX
XX WPI; 1997-479523/44.
XX
XX Conjugate for intracellular delivery - comprising transport moiety
XX having amino acids 49-57 of human immunodeficiency virus tat protein
XX and nucleic acid cargo moiety
XX
XX Disclosure; Columns 55-56; 77pp; English.
XX
PS A novel covalently linked chemical conjugate, comprises a transport
CC polypeptide moiety consisting of at least residues 49-57 of human
CC immunodeficiency virus (HIV) tat protein, but not residues 22-36
CC and 73-86, e.g. the present peptide, and a cargo moiety comprising
CC a single or double stranded nucleic acid. The conjugate can be used
CC to deliver cargo moieties into the cytoplasm and nuclei of cells
CC for therapeutic, prophylactic and diagnostic applications. The HIV
CC tat protein is readily taken up into cells and the cell nucleus.
CC The reduced size of the transport polypeptides minimises
CC interference with the biological activity of the cargo molecule.
CC In addition, by virtue of the absence of the cysteine rich region
CC of the HIV tat protein, the transport polypeptides solve the
CC potential problems of spurious transactivation and disulphide
CC aggregation.
XX
XX Sequence 24 AA:
SQ

```

Query Match 38.2%; Score 55; DB 18; Length 24;  
 Best Local Similarity 84.6%; Pred. No. 0.73;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 YEKRRRRRRRSG 13
   1 | | | | | | | | |
Db 10 YGRKKRRGRRTPG 22

```

RESULT 14  
 AAM26446  
 ID AAM26446 standard; Protein; 24 AA.  
 AC AAM26446;  
 AC  
 DT 16-DEC-1997 (first entry)  
 DT  
 XX HIV tat protein fragment (aa37-58GCC).  
 DE  
 XX  
 KW HIV; tat protein; transport protein; cargo delivery.  
 KW  
 XX Chimeric human immunodeficiency virus type 1;  
 OS Chimeric synthetic.  
 OS  
 PN US5652122-A.  
 PN  
 PD 29-JUL-1997.  
 PD  
 XX 21-DEC-1989; 89US-0454450.  
 XX  
 XX 28-APR-1994; 94US-0235403.  
 PR 21-DEC-1989; 89US-0454450.  
 PR 02-JAN-1991; 91US-0636662.  
 PR 21-AUG-1992; 92US-0934375.  
 PR 19-AUG-1993; 93WO-US07833.  
 PR 24-NOV-1993; 93US-0158015.  
 PR 25-MAY-1995; 95US-0450257.

```

XX (BARS/) BARSOUM J G.
PA (FAME/) FAWELL S E.
PA (FRAN/) FRANKEL A.
PA (PABO/) PABO C.
PA (PEPI/) PEPINSKY R B.
XX Barsoum JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;
XX WPI: 1997-392943/36.
XX
XX New DNA constructs for transporting molecules to cells - encode a
PT fusion protein comprising a modified HIV tat protein and a
PT carboxy-terminal cargo moiety
XX
XX Disclosure: Column 55-56; 76pp; English.
XX
XX This polypeptide comprises amino acids 37-58 of the HIV tat
CC protein (see also AAM26443), plus a Gly-Gly-Cys C-terminal extension.
CC Novel transport proteins comprise modified HIV tat covalently
CC attached to a cargo molecule (see AAM26436-42). The modified tat
CC protein (see AAM26444-49) is characterised by the presence of the tat
CC basic region (amino acids 47-59), and the absence of the Cys-rich
CC region (amino acids 22-36) and the exon 2-encoded C-terminal region
CC (amino acids 73-86). These modifications solve the potential
CC problems of spurious trans-activation and disulphide aggregation,
CC while the reduced size of the transport protein minimises
CC interference with the biological activity of the cargo molecule.
CC DNA molecules that encode the modified tat fusion proteins are
CC claimed and can be used to deliver polypeptides or nucleic acids to
CC the cytoplasm or nuclei of cells in vivo or in vitro.
XX
XX Sequence 24 AA:
SQ

```

```

Query Match 38.2%; Score 55; DB 18; Length 24;
Best Local Similarity 84.6%; Pred. No. 0.73;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 YERKKRRQRRRG 13
   | | | | | | | | |
DB 10 YGRKKRRQRRRG 22

```

```

RESULT 15
AAM76151
ID AAM76151 standard; protein; 24 AA.
XX
XX AAM76151;
XX
XX 24-NOV-1998 (first entry)
XX
XX HIV Type I Tat protein fragment aa. 38-58GCC.
XX
XX Tat protein; cargo molecule; therapy; diagnosis; transport protein;
XX fusion protein; human papillomavirus E2 repressor; target cell.
XX
XX Human immunodeficiency virus type 1.
XX
XX US5804604-A.
XX
XX 08-SEP-1998.
XX
XX 25-MAY-1995; 95US-0450236.
XX
XX 28-APR-1994; 94US-0235403.
XX 21-DEC-1989; 89US-0454450.
XX 02-JAN-1991; 91US-0636662.
XX 19-AUG-1993; 93MO-US07833.
XX 24-NOV-1993; 93US-0158013.
XX 25-MAY-1995; 95US-0450236.
XX
XX (BIOL ) BIOGEN INC.

```

```

XX Barsoum JG, Fawell SE, Frankel A, Pabo C, Pepinsky RB;
XX WPI: 1998-505702/43.
XX
XX HIV tat-derived transport fusion proteins - used to deliver
PT biological active molecules e.g. peptide(s) or nucleic acids,
PT specifically into cytoplasm or nuclei of cells
XX
XX Disclosure: Column 55-56; 83pp; English.
XX
XX This sequence represents a human immunodeficiency virus (HIV) Type I
CC Tat protein fragment, tat38-58GCC which is used in a method for the
CC delivery of biologically active cargo molecules into the cytoplasm and
CC nuclei of cells, for therapeutic, prophylactic or diagnostic purposes.
CC This is accomplished by the presence of a small, basic section of tat
CC transport protein of HIV. This is used as it is this protein which is
CC observed to cause human cells in culture to take up HIV. The method
CC involves the use of a cargo moiety in combination with a transport
CC moiety usually in the form of a fusion protein. The cargo moiety is a
CC human papillomavirus E2 repressor that retains its biological activity
CC after delivery into a target cell and where the transport moiety is
CC one of following HIV tat protein fragments (a) aa 47-58, (b) aa 47-72,
CC (c) 38-72, (d) aa 38-58, (e) aa 37-58, (f) aa 1-21 and 38-72.
CC (g) aa 47-62 or aa 38-62. The proteins allow delivery of specific
CC peptides into cells at high concentrations due to use of existing
CC transporters. Previous methods of delivery include bombardment and
CC transforming, which only allow a fraction of the cell population to be
CC infected and can additionally damage cells as they cause physical opening
CC of the cell walls/membranes to allow entry.
XX
XX Sequence 24 AA:
SQ

```

```

Query Match 38.2%; Score 55; DB 19; Length 24;
Best Local Similarity 84.6%; Pred. No. 0.73;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 YERKKRRQRRRG 13
   | | | | | | | | |
DB 10 YGRKKRRQRRRG 22

```

Search completed: July 15, 2002, 13:51:41  
Job time: 230 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:52:08 ; Search time 21.58 Seconds  
(Without alignments)  
32.824 Million cell updates/sec

Title: US-09-712-819A-10  
Perfect score: 144  
Sequence: 1 YEKRRRRRRSGSGTDTAPASSLQARD 29

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 142465

Minimum DB seq length: 0  
Maximum DB seq length: 29

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCPUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1a1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	56	38.9	18	5	PCT-US95-06077-3
2	56	38.9	26	5	PCT-US95-06077-24
3	55	38.2	15	1	US-08-450-237-6
4	55	38.2	15	1	US-08-450-246-6
5	55	38.2	15	1	US-08-450-098-6
6	55	38.2	15	1	US-08-451-233-6
7	55	38.2	15	1	US-08-450-236-6
8	55	38.2	15	4	US-08-235-403-6
9	55	38.2	24	1	US-08-450-257-4
10	55	38.2	24	1	US-08-450-246-4
11	55	38.2	24	1	US-08-450-098-4
12	55	38.2	24	1	US-08-451-233-4
13	55	38.2	24	1	US-08-450-236-4
14	55	38.2	24	4	US-08-235-403-4
15	55	38.2	25	5	PCT-US91-02942-42
16	55	38.2	26	1	US-08-450-257-48
17	55	38.2	26	1	US-08-450-246-48
18	55	38.2	26	1	US-08-450-098-48
19	55	38.2	26	1	US-08-451-233-48
20	55	38.2	26	1	US-08-450-236-48
21	55	38.2	26	4	US-08-235-403-48
22	54	37.5	25	1	US-08-450-257-53
23	54	37.5	25	1	US-08-450-246-53
24	54	37.5	25	1	US-08-450-098-53
25	54	37.5	25	1	US-08-451-233-53
26	54	37.5	25	1	US-08-450-236-53
27	54	37.5	25	4	US-08-235-403-53

28	54	37.5	26	6	5169933-2	Patent No. 5169933
29	54	37.5	29	2	US-08-928-958-9	Sequence 9, Appl
30	54	37.5	29	2	US-09-072-429-9	Sequence 9, Appl
31	54	37.5	29	4	US-09-111-706-5	Sequence 5, Appl
32	53	36.8	28	4	US-09-041-886-47	Sequence 47, Appl
33	51	35.4	14	5	PCT-US95-06077-4	Sequence 4, Appl
34	51	35.4	14	5	PCT-US95-06077-10	Sequence 10, Appl
35	51	35.4	14	5	PCT-US95-06077-12	Sequence 12, Appl
36	51	35.4	25	5	PCT-US91-02942-26	Sequence 26, Appl
37	51	35.4	28	4	US-09-041-886-40	Sequence 40, Appl
38	50	34.7	11	2	US-08-706-741B-54	Sequence 54, Appl
39	50	34.7	11	2	US-08-924-695A-54	Sequence 54, Appl
40	50	34.7	11	4	US-09-208-966-2	Sequence 2, Appl
41	50	34.7	11	4	US-09-296-089-37	Sequence 37, Appl
42	50	34.7	12	1	US-08-450-257-47	Sequence 47, Appl
43	50	34.7	12	1	US-08-450-246-47	Sequence 47, Appl
44	50	34.7	12	1	US-08-450-098-47	Sequence 47, Appl
45	50	34.7	12	1	US-08-451-233-47	Sequence 47, Appl

## ALIGNMENTS

```
RESULT 1
PCT-US95-06077-3
; Sequence 3, Application PC/TUS9506077
; GENERAL INFORMATION:
; APPLICANT: Immunobiology Research, Institute Inc.
; TITLE OF INVENTION: Vaccine interfection of Extracellular
; TITLE OF INVENTION: Transactivating Proteins of Human Immunodeficiency Virus
; TITLE OF INVENTION: and Other Chronically Infecting Viruses Employing Similar
; TITLE OF INVENTION: InterCellular Transactivating Strategies
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Howson and Howson
; STREET: Spring House Corporate Cntr, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06077
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/247,991
; FILING DATE: 23-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: IRI44PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; PCT-US95-06077-3

Query Match 38.9%, Score 56; DB 5; Length 18;
Best Local Similarity 68.8%; Pred. No. 0.058;
Matches 11; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

DB 2 YGRRKRRQRRRAPOGS 17

RESULT 2  
PCT-US95-06077-24

Sequence 24, Application PC/TUS9506077

GENERAL INFORMATION:

APPLICANT: Immunobiology Research, Institute Inc.

TITLE OF INVENTION: Vaccine Interdiction of Extracellular

TITLE OF INVENTION: Transactivating Proteins of Human Immunodeficiency Virus

TITLE OF INVENTION: and Other Chronically Infecting Viruses Employing Similar

TITLE OF INVENTION: Interleukin Transactivating Strategies

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Cntr, PO Box 457

CITY: Spring House

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/06077

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/247,991

FILING DATE: 23-MAY-1994

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: IR144PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:

LENGTH: 26 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: Peptide

PCT-US95-06077-24

Query Match 38.9%; Score 56; DB 5; Length 26;

Best Local Similarity 68.8%; Pred. No. 0.087;

Matches 11; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YERKKRRQRRRSGCT 16

DB 2 YGRRKRRQRRRAPOGS 17

RESULT 3

US-08-450-257-6

Sequence 6, Application US/08450257

Patent No. 5652122

GENERAL INFORMATION:

APPLICANT: FRANKEL, Alan

APPLICANT: PABO, Carl

APPLICANT: BARSOUM, James G.

APPLICANT: FAMELL, Stephen E.

APPLICANT: PEPINSKY, R. B.

TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES

NUMBER OF SEQUENCES: 69

CORRESPONDENCE ADDRESS:

ADDRESSEE: FISH & NEAVE

STREET: 1251 Avenue of the Americas

CITY: New York  
STATE: New York  
COUNTRY: USA

ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/450,257

FILING DATE: 23-MAY-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/235,403

FILING DATE: 28-APR-1994

APPLICATION NUMBER: US 07/934,375

FILING DATE: 21-AUG-1992

APPLICATION NUMBER: US 07/098,766

FILING DATE: 28-JUL-1993

APPLICATION NUMBER: PCT/US93/07833

FILING DATE: 19-AUG-1993

APPLICATION NUMBER: US 07/454,450

FILING DATE: 21-DEC-1989

APPLICATION NUMBER: US 07/636,662

FILING DATE: 02-JAN-1991

APPLICATION NUMBER: US 08/158,015

FILING DATE: 24-NOV-1993

ATTORNEY/AGENT INFORMATION:

NAME: Haley Jr., James F.

REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: B170 CIP 2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 596-9000

TELEFAX: (212) 596-9090

TELEX: 14-8367

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-450-257-6

Query Match 38.2%; Score 55; DB 1; Length 15;

Best Local Similarity 84.6%; Pred. No. 0.066;

Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YERKKRRQRRRSG 13

DB 1 YGRRKRRQRRRAPOGS 13

RESULT 4

US-08-450-246-6

Sequence 6, Application US/08450246

Patent No. 5670617

GENERAL INFORMATION:

APPLICANT: FRANKEL, Alan

APPLICANT: PABO, Carl

APPLICANT: BARSOUM, James G.

APPLICANT: FAMELL, Stephen E.

TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES

NUMBER OF SEQUENCES: 69

CORRESPONDENCE ADDRESS:

ADDRESSEE: FISH & NEAVE

STREET: 1251 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10020

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,246
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/235,403
; FILING DATE: 28-APR-1994
; APPLICATION NUMBER: US 07/934,375
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: US 07/098,766
; FILING DATE: 28-JUL-1993
; APPLICATION NUMBER: PCT/US93/07833
; FILING DATE: 19-AUG-1993
; APPLICATION NUMBER: US 07/454,450
; FILING DATE: 21-DEC-1989
; APPLICATION NUMBER: US 07/636,662
; FILING DATE: 02-JAN-1991
; APPLICATION NUMBER: US 08/158,015
; FILING DATE: 24-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: B170 CIP 2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; TELEX: 14-8367
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-450-246-6

Query Match      38.2%; Score 55; DB 1; Length 15;
Best Local Similarity 84.6%; Pred. No. 0.066;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 YEKRRRQRRRG 13
DB      1 YGKKRRQRRRG 13

RESULT
US-08-450-098-6
; Sequence 6, Application US/08450098
; Patent No. 5674980
; GENERAL INFORMATION:
; APPLICANT: FRANKEL, Alan
; APPLICANT: PABO, Carl
; APPLICANT: BARSOUM, James G.
; APPLICANT: FAWELL, Stephen E.
; APPLICANT: PEPINSKY, R. B.
; TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,098
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/235,403
; FILING DATE: 28-APR-1994
; APPLICATION NUMBER: US 07/934,375
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: US 07/098,766
; FILING DATE: 28-JUL-1993
; APPLICATION NUMBER: PCT/US93/07833
; FILING DATE: 19-AUG-1993
; APPLICATION NUMBER: US 07/454,450
; FILING DATE: 21-DEC-1989
; APPLICATION NUMBER: US 07/636,662
; FILING DATE: 02-JAN-1991
; APPLICATION NUMBER: US 08/158,015
; FILING DATE: 24-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: B170 CIP 2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; TELEX: 14-8367
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-450-098-6

Query Match      38.2%; Score 55; DB 1; Length 15;
Best Local Similarity 84.6%; Pred. No. 0.066;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 YEKRRRQRRRG 13
DB      1 YGKKRRQRRRG 13

RESULT
US-08-451-233-6
; Sequence 6, Application US/08451233
; Patent No. 5747641
; GENERAL INFORMATION:
; APPLICANT: FRANKEL, Alan
; APPLICANT: PABO, Carl
; APPLICANT: BARSOUM, James G.
; APPLICANT: FAWELL, Stephen E.
; APPLICANT: PEPINSKY, R. B.
; TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/451,233
; FILING DATE: 25-MAY-1995
```

CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/235,403  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993  
APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-451-233-6

Query Match 38.2%; Score 55; DB 1; Length 15;  
Best Local Similarity 84.6%; Pred. No. 0.066;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YEKRRRRRRRG 13  
1 | | | | | | | | | |  
Db 1 YGRKKRRQRRRG 13

RESULT 7  
US-08-450-236-6  
Sequence 6, Application US/08450236  
Patent No. 5804604  
GENERAL INFORMATION:  
APPLICANT: FRANKEL, Alan  
APPLICANT: PABO, Carl  
APPLICANT: BARSOUM, James G.  
APPLICANT: FAMELL, Stephen E.  
APPLICANT: PEPIISKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,236  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/235,403  
FILING DATE: 28-APR-1994

APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993  
APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-450-236-6

Query Match 38.2%; Score 55; DB 1; Length 15;  
Best Local Similarity 84.6%; Pred. No. 0.066;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YEKRRRRRRRG 13  
1 | | | | | | | | | |  
Db 1 YGRKKRRQRRRG 13

RESULT 8  
US-08-235-403-6  
Sequence 6, Application US/08235403  
Patent No. 6316003  
GENERAL INFORMATION:  
APPLICANT: FRANKEL, Alan  
APPLICANT: PABO, Carl  
APPLICANT: BARSOUM, James G.  
APPLICANT: FAMELL, Stephen E.  
APPLICANT: PEPIISKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/235,403  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993



APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-235-403-6

Query Match 38.2%; Score 55; DB 4; Length 15;  
Best Local Similarity 84.6%; Pred. No. 0.066;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YEKRRRRRRSG 13  
| | | | | | | | | | | | | | | | |  
DB 1 YGRRRRRRRRPG 13

RESULT 9  
US-08-450-257-4  
Sequence 4, Application US/08450257  
Patent No. 5652122  
GENERAL INFORMATION:  
APPLICANT: FRANKEL, Alan  
APPLICANT: PABO, Carl  
APPLICANT: BARSDOM, James G.  
APPLICANT: FAMELL, Stephen E.  
APPLICANT: PEPIISKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,257  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/235,403  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993  
APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.

APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-450-257-4

Query Match 38.2%; Score 55; DB 1; Length 24;  
Best Local Similarity 84.6%; Pred. No. 0.11;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YEKRRRRRRSG 13  
| | | | | | | | | | | | | | | | |  
DB 10 YGRRRRRRRRPG 22

RESULT 10  
US-08-450-246-4  
Sequence 4, Application US/08450246  
Patent No. 5670617  
GENERAL INFORMATION:  
APPLICANT: FRANKEL, Alan  
APPLICANT: PABO, Carl  
APPLICANT: BARSDOM, James G.  
APPLICANT: FAMELL, Stephen E.  
APPLICANT: PEPIISKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,246  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/235,403  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993  
APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.

REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-450-098-4

Query Match 38.2%; Score 55; DB 1; Length 24;  
Best Local Similarity 84.6%; Pred. No. 0.11;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YERKKRRRRRG 13  
Db 10 YGKKRRRRRRRG 22

RESULT 11  
US-08-450-098-4  
Sequence 4, Application US/08450098  
Patent No. 5674980  
GENERAL INFORMATION:  
APPLICANT: FRANKEL, Alan  
APPLICANT: BARSOUD, James G.  
APPLICANT: FAWELL, Stephen E.  
APPLICANT: PEPINSKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,098  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/235,403  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993  
APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000

TELEFAX: (212) 596-9090  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-450-098-4

Query Match 38.2%; Score 55; DB 1; Length 24;  
Best Local Similarity 84.6%; Pred. No. 0.11;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YERKKRRRRRG 13  
Db 10 YGKKRRRRRRRG 22

RESULT 12  
US-08-451-233-4  
Sequence 4, Application US/08451233  
Patent No. 5747641  
GENERAL INFORMATION:  
APPLICANT: FRANKEL, Alan  
APPLICANT: BARSOUD, James G.  
APPLICANT: FAWELL, Stephen E.  
APPLICANT: PEPINSKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/451,233  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/235,403  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993  
APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:

LENGTH: 24 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-451-233-4

Query Match 38.2%; Score 55; DB 1; Length 24;  
Best Local Similarity 84.6%; Pred. No. 0.11;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YERKKRRQRRRSG 13  
| | | | | | | | | | | | | | | | | |  
Db 10 YGRKKRRQRRRRPG 22

RESULT 13  
US-08-450-236-4  
Sequence 4, Application US/08450236  
Patent No. 5804604

GENERAL INFORMATION:  
APPLICANT: FRANKEL, Alan  
APPLICANT: PABO, Carl  
APPLICANT: BARSOUM, James G.  
APPLICANT: FAMELL, Stephen E.  
APPLICANT: PERINSKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSER: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,236  
FILING DATE: 25-MAY-1995

CLASSIFICATION: 514  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/235,403  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993  
APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993

ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
TELEX: 14-8367

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

US-08-450-236-4

Query Match 38.2%; Score 55; DB 1; Length 24;  
Best Local Similarity 84.6%; Pred. No. 0.11;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YERKKRRQRRRSG 13  
| | | | | | | | | | | | | | | | | |  
Db 10 YGRKKRRQRRRRPG 22

RESULT 14  
US-08-235-403-4  
Sequence 4, Application US/08235403  
Patent No. 6316003

GENERAL INFORMATION:  
APPLICANT: FRANKEL, Alan  
APPLICANT: PABO, Carl  
APPLICANT: BARSOUM, James G.  
APPLICANT: FAMELL, Stephen E.  
APPLICANT: PERINSKY, R. B.  
TITLE OF INVENTION: TAT-DERIVED TRANSPORT POLYPEPTIDES  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSER: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/235,403  
FILING DATE:

CLASSIFICATION: 514  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/934,375  
FILING DATE: 21-AUG-1992  
APPLICATION NUMBER: US 07/098,766  
FILING DATE: 28-JUL-1993  
APPLICATION NUMBER: PCT/US93/07833  
FILING DATE: 19-AUG-1993  
APPLICATION NUMBER: US 07/454,450  
FILING DATE: 21-DEC-1989  
APPLICATION NUMBER: US 07/636,662  
FILING DATE: 02-JAN-1991  
APPLICATION NUMBER: US 08/158,015  
FILING DATE: 24-NOV-1993

ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: B170 CIP 2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
TELEX: 14-8367

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-235-403-4

Query Match 38.2%; Score 55; DB 4; Length 24;  
Best Local Similarity 84.6%; Pred. No. 0.11;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YEKRRRRRRSG 13  
| | | | | | | | | |  
Db 10 YGRKKRRRRRRPG 22

RESULT 15  
PCT-US91-02942-42  
; Sequence 42, Application PC/TUS9102942  
; GENERAL INFORMATION:  
; APPLICANT: ROTHLEIN, ROBERT  
; APPLICANT: ADALIR, JOHN R  
; APPLICANT: ATHWAL, DILJEET S  
; TITLE OF INVENTION: HUMANIZED CDR-GRAFTED ICAM-1 ANTIBODY  
; NUMBER OF SEQUENCES: 102  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1225 Connecticut Ave, NW Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US91/02942  
; FILING DATE: 19910429  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9009549.8  
; FILING DATE: 27-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FOX, SAM L  
; REGISTRATION NUMBER: 30,353  
; REFERENCE/DOCKET NUMBER: 1011.0586600  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 466-0800  
; TELEFAX: (202) 833-8716  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US91-02942-42

Query Match 38.2%; Score 55; DB 5; Length 25;  
Best Local Similarity 66.7%; Pred. No. 0.12;  
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
OY 12 SGGSTDPATPASSLOAED 29  
| | | | | | | | | |  
Db 2 SGGSTDTLTITSSLOPPD 19

Search completed: July 15, 2002, 13:52:08  
Job time: 152 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 12:52:49 ; Search time 75.67 Seconds  
(without alignments)  
2,260 Million cell updates/sec

Title: US-09-712-819A-1

Perfect score: 32

Sequence: 1 FTITISS 7

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 231628 segs, 24425594 residues

231628

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database :

1: /cgn2\_6/pdata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/pdata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/pdata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/pdata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/pdata/2/1aa/PCRNUS.COMB.pep.\*  
6: /cgn2\_6/pdata/2/1aa/Backfile1.pep.\*

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	100.0	25	5	PCT-US91-02942-26
2	32	100.0	25	5	PCT-US91-02942-42
3	32	100.0	32	2	US-08-470-139-13
4	32	100.0	32	4	US-08-569-147-40
5	32	100.0	32	4	US-09-347-061-13
6	32	100.0	32	4	US-09-425-638A-92
7	32	100.0	32	4	US-09-425-638A-94
8	32	100.0	32	4	US-09-543-004-92
9	32	100.0	32	4	US-09-543-004-94
10	32	100.0	50	5	PCT-US91-02942-8
11	32	100.0	50	5	PCT-US91-02942-9
12	32	100.0	64	2	US-08-765-179B-10
13	32	100.0	64	2	US-08-765-179B-14
14	32	100.0	70	3	US-08-554-840-9
15	32	100.0	76	4	US-08-851-362D-21
16	32	100.0	80	3	US-08-554-840-10
17	32	100.0	80	3	US-08-554-840-12
18	32	100.0	80	3	US-08-554-840-13
19	32	100.0	80	3	US-08-554-840-15
20	32	100.0	93	3	US-08-783-853A-35
21	32	100.0	95	2	US-08-290-592E-19
22	32	100.0	95	2	US-08-290-592E-33
23	32	100.0	95	5	PCT-US95-10053-16
24	32	100.0	95	5	PCT-US95-10053-30
25	32	100.0	95	5	PCT-US96-09448-19
26	32	100.0	95	5	PCT-US96-09448-33
27	32	100.0	96	3	US-08-466-368-6

28	32	100.0	103	4	US-09-240-274-42	Sequence 42, Appl
29	32	100.0	104	1	US-08-276-852-106	Sequence 106, App
30	32	100.0	104	1	US-08-899-575-106	Sequence 106, App
31	32	100.0	104	1	US-08-899-575-106	Sequence 106, App
32	32	100.0	104	5	PCT-US95-08743-106	Sequence 106, App
33	32	100.0	105	1	US-08-276-852-89	Sequence 89, Appl
34	32	100.0	105	1	US-08-899-575-89	Sequence 89, Appl
35	32	100.0	105	1	US-08-899-575-89	Sequence 89, Appl
36	32	100.0	105	5	PCT-US95-08743-89	Sequence 89, Appl
37	32	100.0	106	1	US-07-634-278-7	Sequence 7, Appl1
38	32	100.0	106	1	US-07-634-278-8	Sequence 8, Appl1
39	32	100.0	106	1	US-08-477-728-7	Sequence 7, Appl1
40	32	100.0	106	1	US-08-477-728-8	Sequence 8, Appl1
41	32	100.0	106	1	US-08-276-852-83	Sequence 83, Appl1
42	32	100.0	106	1	US-08-276-852-85	Sequence 85, Appl1
43	32	100.0	106	1	US-08-474-040-7	Sequence 7, Appl1
44	32	100.0	106	1	US-08-474-040-8	Sequence 8, Appl1
45	32	100.0	106	1	US-08-467-200-7	Sequence 7, Appl1

#### ALIGNMENTS

RESULT 1  
PCT-US91-02942-26  
Sequence 26, Application PC/TUS9102942  
GENERAL INFORMATION:  
APPLICANT: ROTHLEIN, ROBERT  
APPLICANT: ADAIR, JOHN R  
TITLE OF INVENTION: HUMANIZED CDR-GRAFTED ICAM-1 ANTIBODY  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Ave. NW Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/02942  
FILING DATE: 19910429  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9009549.8  
FILING DATE: 27-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: FOX, SAM L  
REGISTRATION NUMBER: 30,353  
REFERENCE/DOCKET NUMBER: 1011.0586600  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US91-02942-26

Query Match 100.0%; Score 32; DB 5; Length 25;  
Best local Similarity 100.0%; Pred. No. 1.5; Indels 0;  
Matches 7; Conservative 0; Mismatches 0; Gaps 0;  
Query 1 FTITISS 7  
|||||||

Db 8 FTLTSS 14

## RESULT 2

PCT-US91-02942-42

Sequence 42, Application PC/TUS9102942

GENERAL INFORMATION:

APPLICANT: ROTHLEIN, ROBERT

APPLICANT: ADAIR, JOHN R

APPLICANT: ATWAL, DILJEET S

TITLE OF INVENTION: HUMANIZED CDR-GRAFTED ICAM-1 ANTIBODY

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein &amp; Fox

STREET: 1225 Connecticut Ave. NW Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US91/02942

FILING DATE: 19910429

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9009549.8

FILING DATE: 27-Apr-1990

ATTORNEY/AGENT INFORMATION:

NAME: FOX, SAM L

REGISTRATION NUMBER: 30,353

REFERENCE/DOCKET NUMBER: 1011,0586600

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 466-0800

TELEFAX: (202) 833-8716

INFORMATION FOR SEQ ID NO: 42:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US91-02942-42

Query Match 100.0%; Score 32; DB 5; Length 25;

Best Local Similarity 100.0%; Pred. No. 1.5;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7

Db 8 FTLTSS 14

RESULT 3

US-08-470-139-13

Sequence 13, Application US/08470139

Patent No. 5998586

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Interleukin-5 specific recombinant antibodies

NUMBER OF SEQUENCES: 28

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/470,139

FILING DATE: 06 JUNE-1995

CLASSIFICATION: 536

Query Match 100.0%; Score 32; DB 4; Length 32;

Best Local Similarity 100.0%; Pred. No. 2;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7

Db 15 FTLTSS 21

ATTORNEY/AGENT INFORMATION:

NAME: TRUJILLO, DOREEN YATKO

REGISTRATION NUMBER: 35,719

REFERENCE/DOCKET NUMBER: CARP-0044

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 32 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-470-139-13

Query Match 100.0%; Score 32; DB 2; Length 32;

Best Local Similarity 100.0%; Pred. No. 2;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7

Db 15 FTLTSS 21

RESULT 4

US-08-569-147-40

Sequence 40, Application US/08569147

Patent No. 6180377

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: HUMANISED ANTIBODIES

NUMBER OF SEQUENCES: 95

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &amp;

STREET: One Liberty Place - 46th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: U.S.A.

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/569,147

FILING DATE: 25-March-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Trujillo, Doreen Yatko

REGISTRATION NUMBER: 35,719

REFERENCE/DOCKET NUMBER: CARP-0047

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-3100

TELEFAX: (215) 568-3439

INFORMATION FOR SEQ ID NO: 40:

SEQUENCE CHARACTERISTICS:

LENGTH: 32 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-569-147-40

Query Match 100.0%; Score 32; DB 4; Length 32;

Best Local Similarity 100.0%; Pred. No. 2;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7

Db 15 FTLTSS 21

```
RESULT 5
US-09-347-061-13
; Sequence 13, Application US/09347061
; Patent No. 6316227
; GENERAL INFORMATION:
; APPLICANT: Bodmer, Mark
; APPLICANT: Athwal, Diljeet Singh
; APPLICANT: Entlage, John Spencer
; TITLE OF INVENTION: Interleukin-5 Specific Recombinant Antibodies
; FILE REFERENCE: CARP-0071
; CURRENT APPLICATION NUMBER: US/09/347,061
; CURRENT FILING DATE: 1999-07-02
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Consensus
US-09-347-061-13
```

```
Query Match          100.0%; Score 32; DB 4; Length 32;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 FTLTSS 7
    |||||
Db 15 FTLTSS 21
```

```
RESULT 6
US-09-425-638A-92
; Sequence 92, Application US/09425638A
; Patent No. 6342587
; GENERAL INFORMATION:
; APPLICANT: Carlos F. Barbas III, Christoph Rader, Gerd Rittler, Sydney Welt and
; APPLICANT: Lloyd J. Old
; TITLE OF INVENTION: A33 ANTIGEN SPECIFIC IMMUNOGLOBULIN PRODUCTS AND USES THEREOF
; FILE REFERENCE: LUD 5630
; CURRENT APPLICATION NUMBER: US/09/425,638A
; CURRENT FILING DATE: 1999-10-22
; NUMBER OF SEQ ID NOS: 129
; SEQ ID NO 92
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-425-638A-92
```

```
Query Match          100.0%; Score 32; DB 4; Length 32;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 FTLTSS 7
    |||||
Db 15 FTLTSS 21
```

```
RESULT 7
US-09-425-638A-94
; Sequence 94, Application US/09425638A
; Patent No. 6342587
; GENERAL INFORMATION:
; APPLICANT: Carlos F. Barbas III, Christoph Rader, Gerd Rittler, Sydney Welt and
; APPLICANT: Lloyd J. Old
; TITLE OF INVENTION: A33 ANTIGEN SPECIFIC IMMUNOGLOBULIN PRODUCTS AND USES THEREOF
; FILE REFERENCE: LUD 5630
; CURRENT APPLICATION NUMBER: US/09/425,638A
```

```
; CURRENT FILING DATE: 1999-10-22
; NUMBER OF SEQ ID NOS: 129
; SEQ ID NO 94
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-425-638A-94
```

```
Query Match          100.0%; Score 32; DB 4; Length 32;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 FTLTSS 7
    |||||
Db 15 FTLTSS 21
```

```
RESULT 8
US-09-543-004-92
; Sequence 92, Application US/09543004
; Patent No. 6346249
; GENERAL INFORMATION:
; APPLICANT: Carlos F. Barbas III, Christoph Rader, Gerd Rittler, Sydney Welt and
; APPLICANT: Lloyd J. Old
; TITLE OF INVENTION: A33 ANTIGEN SPECIFIC IMMUNOGLOBULIN PRODUCTS AND USES THEREOF
; FILE REFERENCE: LUD 5630.1
; CURRENT APPLICATION NUMBER: US/09/543,004
; CURRENT FILING DATE: 2000-04-04
; PRIOR FILING DATE: 09/425,638
; NUMBER OF SEQ ID NOS: 129
; SEQ ID NO 92
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-543-004-92
```

```
Query Match          100.0%; Score 32; DB 4; Length 32;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 FTLTSS 7
    |||||
Db 15 FTLTSS 21
```

```
RESULT 9
US-09-543-004-94
; Sequence 94, Application US/09543004
; Patent No. 6346249
; GENERAL INFORMATION:
; APPLICANT: Carlos F. Barbas III, Christoph Rader, Gerd Rittler, Sydney Welt and
; APPLICANT: Lloyd J. Old
; TITLE OF INVENTION: A33 ANTIGEN SPECIFIC IMMUNOGLOBULIN PRODUCTS AND USES THEREOF
; FILE REFERENCE: LUD 5630.1
; CURRENT APPLICATION NUMBER: US/09/543,004
; CURRENT FILING DATE: 2000-04-04
; PRIOR FILING DATE: 09/425,638
; NUMBER OF SEQ ID NOS: 129
; SEQ ID NO 94
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-543-004-94
```

```
Query Match          100.0%; Score 32; DB 4; Length 32;
```

Best Local Similarity 100.0%; Pred. No. 2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||||  
Db 15 FTLTSS 21

RESULT 10  
PCT-US91-02942-8

Sequence 8, Application PC/TUS9102942  
GENERAL INFORMATION:  
APPLICANT: ROTHLEIN, ROBERT  
APPLICANT: ADAIR, JOHN R  
APPLICANT: ATHMAL, DILJEET S  
TITLE OF INVENTION: HUMANIZED CDR-GRAFTED ICAM-1 ANTIBODY  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Ave. NW Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/02942  
FILING DATE: 19910429  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9009549.8  
FILING DATE: 27-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: FOX, SAM L  
REGISTRATION NUMBER: 30,353  
REFERENCE/DOCKET NUMBER: 1011.0586600  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ. ID NO.: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: Peptide  
PCT-US91-02942-8

Query Match 100.0%; Score 32; DB 5; Length 50;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||||  
Db 26 FTLTSS 32

RESULT 11  
PCT-US91-02942-9

Sequence 9, Application PC/TUS9102942  
GENERAL INFORMATION:  
APPLICANT: ROTHLEIN, ROBERT  
APPLICANT: ADAIR, JOHN R  
APPLICANT: ATHMAL, DILJEET S  
TITLE OF INVENTION: HUMANIZED CDR-GRAFTED ICAM-1 ANTIBODY  
NUMBER OF SEQUENCES: 102  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Ave. NW Suite 300

CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/02942  
FILING DATE: 19910429  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9009549.8  
FILING DATE: 27-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: FOX, SAM L  
REGISTRATION NUMBER: 30,353  
REFERENCE/DOCKET NUMBER: 1011.0586600  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ. ID NO.: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US91-02942-9

Query Match 100.0%; Score 32; DB 5; Length 50;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||||  
Db 26 FTLTSS 32

RESULT 12  
US-08-765-179B-10  
Sequence 10, Application US/08765179B  
Patent No. 5854027  
GENERAL INFORMATION:  
APPLICANT: STEINBE, Boris  
APPLICANT: STEINBACHER, Stefan  
TITLE OF INVENTION: PROCESS FOR MODIFYING THE STABILITY  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP  
STREET: 655 Fifteenth Street N.W. Suite 330  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-5701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/765,179B  
FILING DATE: 14-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/02626  
FILING DATE: 06-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 25 115.7  
FILING DATE: 15-JUL-1994



INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 64 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-765-179B-10

## Query Match

Best Local Similarity 100.0%; Score 32; DB 2; Length 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7  
|||||  
DB 40 FTLTSS 46

## RESULT 13

US-08-765-179B-14  
Sequence 14, Application US/08765179B  
Patent No. 5854027  
GENERAL INFORMATION:  
APPLICANT: STEIDE, Boris  
APPLICANT: STEINBACHER, Stefan  
TITLE OF INVENTION: PROCESS FOR MODIFYING THE STABILITY  
TITLE OF INVENTION: OF ANTIBODIES  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Mikaido, Marmelstein, Murray & Oram LLP  
STREET: 655 Fifteenth Street N.W. Suite 330  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-5701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/765,179B  
FILING DATE: 14-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/02626  
FILING DATE: 06-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 25 115.7  
FILING DATE: 15-JUL-1994  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 64 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-765-179B-14

Query Match  
Best Local Similarity 100.0%; Score 32; DB 2; Length 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7  
|||||  
DB 40 FTLTSS 46

RESULT 14  
US-08-554-840-9  
Sequence 9, Application US/08554840

Patent No. 6001358  
GENERAL INFORMATION:  
APPLICANT: BLACK, Amelia  
APPLICANT: HANNA, Nabil  
APPLICANT: PADLAN, Eduardo A.  
APPLICANT: NEWMAN, Roland A.  
TITLE OF INVENTION: HUMANIZED ANTIBODIES TO HUMAN gp39,  
COMPOSITIONS CONTAINING AND THERAPEUTIC USE THEREOF  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/554,840  
FILING DATE: 07-NOV-1995  
CLASSIFICATION: 336  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 012712-127  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 70 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-554-840-9

Query Match  
Best Local Similarity 100.0%; Score 32; DB 3; Length 70;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7  
|||||  
DB 53 FTLTSS 59

RESULT 15  
US-08-851-362D-21  
Sequence 21, Application US/08851362D  
Patent No. 6235863  
GENERAL INFORMATION:  
APPLICANT: Jakobovits, Aya  
APPLICANT: Yang, Xiao-Dong  
APPLICANT: Gallo, Michael  
APPLICANT: Jia, Xiao-Chi  
TITLE OF INVENTION: Human Monoclonal Antibodies to Epidermal  
FILE REFERENCE: Cell 4.20  
CURRENT APPLICATION NUMBER: US/08/851,362D  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 21  
LENGTH: 76  
TYPE: PRT  
ORGANISM: human  
US-08-851-362D-21

Query Match 100.0%; Score 32; DB 4; Length 76;  
Best Local Similarity 100.0%; Pred. No. 4.9;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7  
|||||||  
Db 52 FTLTSS 58

Search completed: July 15, 2002, 12:59:16  
Job time: 387 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 15, 2002, 12:53:14 ; Search time 95.45 seconds  
(without alignments)  
7.047 Million cell updates/sec

Title: US-09-712-819A-1  
Perfect score: 32  
Sequence: 1 FTLTSS 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	32	100.0	54	2 JT0521	Ig kappa chain V-1
2	32	100.0	71	2 S21526	Ig kappa chain V r
3	32	100.0	77	2 D30502	Ig kappa chain V r
4	32	100.0	77	2 S70443	Ig kappa chain V r
5	32	100.0	78	2 S34102	Ig kappa chain V r
6	32	100.0	79	2 S24215	Ig kappa chain - m
7	32	100.0	81	2 PH1048	Ig light chain V r
8	32	100.0	83	2 S78489	Ig kappa chain V r
9	32	100.0	84	2 S34099	Ig kappa chain V r
10	32	100.0	86	2 S16826	Ig kappa chain V r
11	32	100.0	86	2 S16834	Ig kappa chain V r
12	32	100.0	86	2 S16836	Ig kappa chain V r
13	32	100.0	86	2 S34086	Ig kappa chain V r
14	32	100.0	86	2 S16840	Ig kappa chain V r
15	32	100.0	86	2 S16837	Ig kappa chain V r
16	32	100.0	86	2 S16833	Ig kappa chain V-1
17	32	100.0	86	2 S16830	Ig kappa chain V r
18	32	100.0	86	2 S16824	Ig kappa chain V r
19	32	100.0	86	2 S16829	Ig kappa chain V r
20	32	100.0	87	2 S34098	Ig kappa chain V r
21	32	100.0	87	2 S21523	Ig kappa chain V r
22	32	100.0	87	2 S34097	Ig kappa chain V r
23	32	100.0	87	2 S34084	Ig kappa chain V r
24	32	100.0	87	2 S34083	Ig kappa chain V r
25	32	100.0	88	2 S21528	Ig kappa chain V r
26	32	100.0	88	2 S21525	Ig kappa chain V r
27	32	100.0	88	2 S21520	Ig kappa chain V r
28	32	100.0	88	2 S21522	Ig kappa chain V r
29	32	100.0	88	2 S34104	Ig kappa chain V r

30	32	100.0	91	2 PH1071	Ig light chain V r
31	32	100.0	91	2 S37525	Ig kappa chain V r
32	32	100.0	91	2 S37527	Ig kappa chain V r
33	32	100.0	91	2 S37521	Ig kappa chain V r
34	32	100.0	91	2 S37511	Ig kappa chain V r
35	32	100.0	91	2 S37515	Ig kappa chain V r
36	32	100.0	92	2 S37533	Ig kappa chain V r
37	32	100.0	92	2 S37530	Ig kappa chain V r
38	32	100.0	92	2 S37529	Ig kappa chain V r
39	32	100.0	92	2 S37523	Ig kappa chain V r
40	32	100.0	92	2 S37535	Ig kappa chain V r
41	32	100.0	92	2 S37509	Ig kappa chain V r
42	32	100.0	92	2 S37512	Ig kappa chain V r
43	32	100.0	92	2 S37531	Ig kappa chain V r
44	32	100.0	92	2 S37534	Ig kappa chain V r
45	32	100.0	92	2 S37532	Ig kappa chain V r

## ALIGNMENTS

RESULT 1  
JT0521  
Ig kappa chain V-II region (Cp1) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 23-Oct-1992 #sequence\_revision 23-Oct-1992 #text\_change 09-May-1997  
C:Accession: JT0521  
R:Anker, R.; Conley, M.E.; Pollak, B.A.  
J. Exp. Med. 169, 2109-2119, 1989  
A:Title: Clonal diversity in the B cell repertoire of patients with X-linked agammaglobulinemia  
A:Reference number: JT0511; MUID:89279157  
A:Accession: JT0521  
A:Molecule type: mRNA  
A:Residues: 1-54 <ANK>  
A:Note: the sequence shown here is one of eight productive V-D-J mu chain rearrangements  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotrimer; Immunoglobulin  
F:1-49/Domain: V region <VRE>  
F:50-54/Domain: J region <JRE>

Query Match 100.0%; Score 32; DB 2; Length 54;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||  
DB 23 FTLTSS 29

RESULT 2  
S21526  
Ig kappa chain V region - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 20-Feb-1995 #sequence\_revision 25-Oct-1996 #text\_change 23-Jul-1999  
C:Accession: S34082; S21526  
R:Wagner, S.D.; Juzatito, L.  
Eur. J. Immunol. 23, 391-397, 1993  
A:Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distributed  
A:Reference number: S34076; MUID:93170387  
A:Accession: S34082  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-71 <MA2>  
A:Cross-references: EMBL:X66042; NID:g33318; PDB:CAA6841.1; PTD:g33319  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotrimer; Immunoglobulin

Query Match 100.0%; Score 32; DB 2; Length 71;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||||  
Db 64 FTLTSS 70

## RESULT 3

D30502  
Ig kappa chain V region (D44) - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 03-Nov-1988 #sequence\_revision 03-Aug-1992 #text\_change 23-Jul-1999  
C:Accession: D30502  
R:Ellat, D.; Webster, D.M.; Rees, A.R.  
J. Immunol. 141, 1745-1753, 1988  
A:Title: V region sequences of anti-DNA and anti-RNA autoantibodies from NZB/NZW F-1 M1C  
A:Reference number: A30502; MUID:88315787  
A:Accession: D30502  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-77 <ELT>  
A:Cross-references: GB:M21908; NID:q197073; PIDN:AAA8908.1; PID:q197074  
A:Note: the authors translated the codon CAG for residue 48 as Pro  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 32; DB 2; Length 77;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||||  
Db 40 FTLTSS 46

## RESULT 4

S70443  
Ig kappa chain V region - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 07-May-1999  
C:Accession: S70443  
R:Cuisinier, A.M.; Fumoux, F.; Fougereau, M.; Tonnelle, C.  
Mol. Immunol. 29, 1363-1373, 1992  
A:Title: IGM kappa/lambda EBV human B cell clone: an early step of differentiation of f  
A:Reference number: S70442; MUID:93024508  
A:Accession: S70443  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-77 <CU1>  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology

Query Match 100.0%; Score 32; DB 2; Length 77;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||||  
Db 40 FTLTSS 46

## RESULT 5

S34102  
Ig kappa chain V region - human  
C:Species: Homo sapiens (man)  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 07-May-1999  
C:Accession: S34102  
R:Wagner, S.D.; Luzzatto, L.  
Eur. J. Immunol. 23, 391-397, 1993  
A:Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distributed  
A:Reference number: S34076; MUID:93170387  
A:Accession: S34102  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-78 <WAG>  
A:Cross-references: EMBL:X67186  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 32; DB 2; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||||  
Db 62 FTLTSS 68

## RESULT 6

S24215  
Ig kappa chain - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999  
C:Accession: S24215  
R:Shimizu, T.; Iwasato, T.; Yamagishi, H.  
J. Exp. Med. 173, 1065-1072, 1991  
A:Title: Deletions of immunoglobulin C(kappa) region characterized by the circular ex  
A:Reference number: S24214; MUID:91217618  
A:Accession: S24215  
A:Status: preliminary; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-79 <SHI>  
A:Cross-references: EMBL:X58202; NID:q53718; PIDN:CAA41178.1; PID:q930195  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 32; DB 2; Length 79;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||||  
Db 44 FTLTSS 50

## RESULT 7

PH1048  
Ig light chain V region (clone 165.49) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 15-Jun-1996  
C:Accession: PH1048  
R:Fillman, D.M.; Jou, N.T.; Hill, R.J.; Marion, T.N.  
J. Exp. Med. 176, 761-779, 1992  
A:Title: Both Igm and Igg anti-DNA antibodies are the products of clonally selective  
A:Reference number: PH0971; MUID:92381444  
A:Accession: PH1048  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-81 <TID>  
A:Experimental source: B cell, strain [NZB x NZW]F1  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: immunoglobulin

Query Match 100.0%; Score 32; DB 2; Length 81;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||||  
Db 56 FTLTSS 62

## RESULT 8

S78489

Ig kappa chain V region (patient 28) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 30-Apr-1998 #sequence\_revision 08-May-1998 #text\_change 21-Jan-2000  
C:Accession: S78489; S34100  
R:Wagner, S.  
submitted to the EMBL Data Library, July 1992  
A:Reference number: S78488  
A:Accession: S78489  
A:Molecule type: DNA  
A:Residues: 1-83 <NAG>  
A:Cross-references: EMBL:X67184  
A:Experimental source: Patient 28  
R:Wagner, S.D.; Luzzatto, L.  
Eur. J. Immunol. 23, 391-397, 1993  
A:Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distributed  
A:Reference number: S34076; MUID:93170387  
A:Accession: S34100  
A>Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-68, 'Q', 70-73, 'A', 75-83 <NAG>  
A:Cross-references: EMBL:X67184  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
P:5-80/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 32; DB 2; Length 83;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||  
DB 61 FTLTSS 67

RESULT 9  
S34099  
Ig kappa chain V region - human  
C:Species: Homo sapiens (man)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 21-Jan-2000  
C:Accession: S34099  
R:Wagner, S.D.; Luzzatto, L.  
Eur. J. Immunol. 23, 391-397, 1993  
A:Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distributed  
A:Reference number: S34076; MUID:93170387  
A:Accession: S34099  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-84 <NAG>  
A:Cross-references: EMBL:X67183  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
P:5-79/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 32; DB 2; Length 84;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||  
DB 60 FTLTSS 66

RESULT 10  
S16826  
Ig kappa chain V region - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000  
C:Accession: S16826; S34101  
R:Blaison, G.; Kuntz, J.L.; Pasquali, J.L.  
Eur. J. Immunol. 21, 1221-1227, 1991  
A:Title: Molecular analysis of V(kappa)III variable regions of polyclonal rheumatoid

A:Reference number: S16823; MUID:91243737  
A:Accession: S16826  
A>Status: preliminary; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-86 <BLA>  
A:Cross-references: EMBL:X54824; NID:q33653; PIDN:CAA38593.1; PID:q33654  
R:Wagner, S.D.; Luzzatto, L.  
Eur. J. Immunol. 23, 391-397, 1993  
A:Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distrib  
A:Reference number: S34076; MUID:93170387  
A:Accession: S34101  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-86 <NAG>  
A:Cross-references: EMBL:X67185  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
P:7-81/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 32; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||  
DB 62 FTLTSS 68

RESULT 11  
S16834  
Ig kappa chain V region - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000  
C:Accession: S16834  
R:Blaison, G.; Kuntz, J.L.; Pasquali, J.L.  
Eur. J. Immunol. 21, 1221-1227, 1991  
A:Title: Molecular analysis of V(kappa)III variable regions of polyclonal rheumatoid  
A:Reference number: S16823; MUID:91243737  
A:Accession: S16834  
A>Status: preliminary; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-86 <BLA>  
A:Cross-references: EMBL:X54832  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
P:7-81/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 32; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||  
DB 62 FTLTSS 68

RESULT 12  
S16836  
Ig kappa chain V region - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000  
C:Accession: S16836  
R:Blaison, G.; Kuntz, J.L.; Pasquali, J.L.  
Eur. J. Immunol. 21, 1221-1227, 1991  
A:Title: Molecular analysis of V(kappa)III variable regions of polyclonal rheumatoid  
A:Reference number: S16823; MUID:91243737  
A:Accession: S16836  
A>Status: preliminary; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-86 <BLA>  
A:Cross-references: EMBL:X54834

C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin  
F:7-81/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 32; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||||  
DB 62 FTLTSS 68

## RESULT 13

S34086  
Ig kappa chain V region - human  
C:Species: Homo sapiens (man)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 21-Jan-2000  
C:Accession: S34086  
R:Blaisson, G.; Kuntz, J.L.; Pasquali, J.L.  
Eur. J. Immunol. 23, 391-397, 1993  
A:Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distribute  
A:Reference number: S34076; MUID:93170387  
A:Accession: S34086  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-86 <WAG>  
A:Cross-references: EMBL:X67169  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin  
F:9-83/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 32; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||||  
DB 64 FTLTSS 70

## RESULT 14

S16840  
Ig kappa chain V region - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000  
C:Accession: S16840  
R:Blaisson, G.; Kuntz, J.L.; Pasquali, J.L.  
Eur. J. Immunol. 21, 1221-1227, 1991  
A:Title: Molecular analysis of V(kappa)III variable regions of polyclonal rheumatoid fac  
A:Reference number: S16823; MUID:91243737  
A:Accession: S16840  
A>Status: preliminary; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-86 <BLA>  
A:Cross-references: EMBL:X54838  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin  
F:7-81/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 32; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||||  
DB 62 FTLTSS 68

## RESULT 15

S16837  
Ig kappa chain V region - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000  
C:Accession: S16837  
R:Blaisson, G.; Kuntz, J.L.; Pasquali, J.L.  
Eur. J. Immunol. 21, 1221-1227, 1991  
A:Title: Molecular analysis of V(kappa)III variable regions of polyclonal rheumatoid

A:Reference number: S16823; MUID:91243737  
A:Accession: S16837  
A>Status: preliminary; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-86 <BLA>  
A:Cross-references: EMBL:X54835  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin  
F:7-81/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 32; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
|||||||  
DB 62 FTLTSS 68

Search completed: July 15, 2002, 13:00:57  
Job time: 463 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 12:59:24 ; Search time 44.9 Seconds  
(without alignments)  
6.036 Million cell updates/sec

Title: US-09-712-819A-1

Perfect score: 32  
Sequence: 1 FTLTSS 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	32	100.0	107 1	KVID_HUMAN
2	32	100.0	108 1	KVIF_HUMAN
3	32	100.0	108 1	KVIG_HUMAN
4	32	100.0	108 1	KVIE_HUMAN
5	32	100.0	108 1	KVIF_HUMAN
6	32	100.0	108 1	KVIF_HUMAN
7	32	100.0	108 1	KVIF_HUMAN
8	32	100.0	108 1	KVIF_HUMAN
9	32	100.0	108 1	KVIF_HUMAN
10	32	100.0	109 1	KVIF_HUMAN
11	32	100.0	109 1	KVIF_HUMAN
12	32	100.0	114 1	KVIF_HUMAN
13	32	100.0	114 1	KVIF_HUMAN
14	32	100.0	115 1	KVIF_HUMAN
15	32	100.0	116 1	KVIF_HUMAN
16	32	100.0	117 1	KVIF_HUMAN
17	32	100.0	117 1	KVIF_HUMAN
18	32	100.0	121 1	KVIF_HUMAN
19	32	100.0	129 1	KVIF_HUMAN
20	32	100.0	129 1	KVIF_HUMAN
21	32	100.0	133 1	KVIF_HUMAN
22	32	100.0	134 1	KVIF_HUMAN
23	32	100.0	136 1	KVIF_HUMAN
24	32	100.0	136 1	KVIF_HUMAN
25	32	100.0	136 1	KVIF_HUMAN
26	32	100.0	136 1	KVIF_HUMAN
27	32	100.0	136 1	KVIF_HUMAN
28	32	100.0	136 1	KVIF_HUMAN
29	32	100.0	136 1	KVIF_HUMAN
30	32	100.0	136 1	KVIF_HUMAN
31	32	100.0	136 1	KVIF_HUMAN
32	32	100.0	136 1	KVIF_HUMAN
33	32	100.0	136 1	KVIF_HUMAN

34	28	87.5	108 1	KV05_RABIT	P01686 oryctolagus
35	28	87.5	108 1	KV06_RABIT	P01687 oryctolagus
36	28	87.5	108 1	KV07_RABIT	P01688 oryctolagus
37	28	87.5	108 1	KV08_RABIT	P01689 oryctolagus
38	28	87.5	108 1	KV1B_HUMAN	P01594 homo sapien
39	28	87.5	108 1	KV1E_HUMAN	P01597 homo sapien
40	28	87.5	108 1	KV1F_HUMAN	P01603 homo sapien
41	28	87.5	108 1	KV1G_HUMAN	P01605 homo sapien
42	28	87.5	108 1	KV1H_HUMAN	P01606 homo sapien
43	28	87.5	108 1	KV1I_HUMAN	P01608 homo sapien
44	28	87.5	108 1	KV1J_HUMAN	P01609 homo sapien
45	28	87.5	108 1	KV1K_HUMAN	P01619 homo sapien

## ALIGNMENTS

RESULT 1  
KVID\_HUMAN STANDARD: PRT; 107 AA.  
AC P01596;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Ig kappa chain V-I region CAR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=75075135; PubMed=4216454;  
RA Milstein C.P., Deverson E.V.;  
RT "Primary structure of kappa light chain from a human myeloma protein."  
RT Eur. J. Biochem. 49:377-391(1974).  
CC -I- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (1,2) MARKER.  
CC -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN.  
DR PIR: A01864; KIHUAR.  
DR HSSP: P80362; IWT.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003596; Ig\_V.  
DR Pfam: PF00047; Ig\_1.  
DR SMART: SM00406; IgV\_1.  
FT CARBOHYD 28 N-LINKED (GLCNAC. . .).  
FT NON\_TER 107  
SQ SEQUENCE 107 AA; 11703 MW; E1BF0DF9844C3346 CRC64;

Query Match 100.0%; Score 32; DB 1; Length 107;  
Best local similarity 100.0%; Pred. No. 0.8; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7  
DB 71 FTLTSS 77

RESULT 2  
KVIF\_HUMAN STANDARD: PRT; 108 AA.  
AC P01598;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Ig kappa chain V-I region EU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]

```

RP SEQUENCE.
RX MEDLINE=71064023; PubMed=5489770;
RA Goltlieb P.D., Cunningham B.A., Rutishauser U., Eelaman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. VI. Amino
RL Biochemistry 9:3155-3161(1970).
RN (2)
RP DISULFIDE BOND.
RX MEDLINE=71064027; PubMed=4923144;
RA Gall W.E., Eelaman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. X.
RL Intrachain disulfide bonds."
CC Biochemistry 9:3188-3196(1970).
CC -1- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (3) MARKER.
DR PIR: A01866; KIHUHU.
DR HSSP: P01607; IREI.
DR InterPro: IPR003006; Iq_MHC.
DR InterPro: IPR003596; Iq_V.
DR Pfam: PF00047; Iq; 1.
DR SMART: SM00406; IGV; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 23
FT DOMAIN 24 34
FT DOMAIN 35 49
FT DOMAIN 50 56
FT DOMAIN 57 88
FT DOMAIN 89 97
FT DOMAIN 98 107
FT DISULFID 23 88
FT NON_TER 108
SQ SEQUENCE 108 AA; 11788 MW; 9CD294F2P4D88823 CRC64;

Query Match
Best Local Similarity 100.0%; Score 32; DB 1; Length 108;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7
DB 71 FTLTSS 77

RESULT 3
KV1G_HUMAN STANDARD; PRT; 108 AA.
ID KVLG_HUMAN
AC P01593;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Iq kappa chain V-I region GAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=75059122; PubMed=4215718;
RA Laure C.J., Watanabe S., Hilschmann N.;
RT "The primary structure of a monoclonal Igm-immunoglobulin
RT (macroglobulin Gal.), I. The amino acid sequence of the L-chain of
RT kappa-type, subgroup I."
RL Hoppe-Sevler's Z. Physiol. Chem. 354:1503-1504(1973).
CC -1- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (3) MARKER.
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A WALDENSTROM'S
CC MACROGLOBULIN.
DR PIR: A01867; KIHUGL.
DR HSSP: P01607; IREI.
DR InterPro: IPR003006; Iq_MHC.
DR InterPro: IPR003596; Iq_V.
DR Pfam: PF00047; Iq; 1.
DR SMART: SM00406; IGV; 1.
KW Immunoglobulin V region.

```

```

FT DOMAIN 1 23
FT DOMAIN 24 34
FT DOMAIN 35 49
FT DOMAIN 50 56
FT DOMAIN 57 88
FT DOMAIN 89 97
FT DOMAIN 98 107
FT DISULFID 23 88
FT NON_TER 108
SQ SEQUENCE 108 AA; 11814 MW; C1AD3CBDF60DF73 CRC64;

Query Match
Best Local Similarity 100.0%; Score 32; DB 1; Length 108;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7
DB 71 FTLTSS 77

RESULT 4
KV1H_HUMAN STANDARD; PRT; 108 AA.
ID KVLH_HUMAN
AC P01600;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Iq kappa chain V-I region Hau.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=71032830; PubMed=4097974;
RA Watanabe S., Hilschmann N.;
RT "The primary structure of a monoclonal kappa-type immunoglobulin L-
RT chain of subgroup I (Bence-Jones Protein Hau); subdivision within
RT subgroups."
RL Hoppe-Sevler's Z. Physiol. Chem. 351:1291-1295(1970).
CC -1- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (3) MARKER.
CC -1- MISCELLANEOUS: THIS IS A BENCE-JONES PROTEIN.
DR PIR: A01868; KIH0HU.
DR HSSP: P80362; IWTU.
DR InterPro: IPR003006; Iq_MHC.
DR InterPro: IPR003596; Iq_V.
DR Pfam: PF00047; Iq; 1.
DR SMART: SM00406; IGV; 1.
KW Immunoglobulin V region; Bence-Jones protein.
FT DOMAIN 1 23
FT DOMAIN 24 34
FT DOMAIN 35 49
FT DOMAIN 50 56
FT DOMAIN 57 88
FT DOMAIN 97 97
FT DOMAIN 98 107
FT DISULFID 23 88
FT NON_TER 108
SQ SEQUENCE 108 AA; 11671 MW; 08D3A6160D8D0618 CRC64;

Query Match
Best Local Similarity 100.0%; Score 32; DB 1; Length 108;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7
DB 71 FTLTSS 77

RESULT 5
KV1S_HUMAN

```



```

ID KY15_HUMAN STANDARD; PRT; 108 AA.
AC P01611;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-I region Mes.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=81092279; PubMed=6778806;
RA Kratzin H., Yang C., Krusche J.U., Hilschmann N.;
RT *Preparative separation of the tryptic hydrolysate of a protein by
RT high pressure liquid chromatography. The primary structure of a
RT monoclonal L-chain of K-type, subgroup I (Bence-Jones Protein
RT Mes).
RL Hoppe-Seyler's Z. Physiol. Chem. 361:1591-1598(1980).
CC -1- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (3) MARKER.
CC PIR: A01877; KIHMS.
DR HSSP; P80362; 1MTL.
DR InterPro; IPR003596; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IgV; 1.
KW Immunoglobulin V region; Bence-Jones protein.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 2 34 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 3 35 49 FRAMEWORK-2.
FT DOMAIN 4 50 56 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 5 57 88 FRAMEWORK-3.
FT DOMAIN 6 89 97 FRAMEWORK-4.
FT DOMAIN 7 98 107 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 23 88 FRAMEWORK-4.
FT NON_TER 108 BY SIMILARITY.
SQ SEQUENCE 108 AA; 11608 MW; 782B14A649A60E45 CRC64;

Query Match 100.0%; Score 32; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7
DB 71 FTLTSS 77

RESULT 6
KV5Q_MOUSE STANDARD; PRT; 108 AA.
ID KV5Q_MOUSE
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-V region UPC 61.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=79195288; PubMed=109517;
RA Vrana M., Rudikoff S., Potter M.;
RT *The structural basis of a hapten-inhibitable kappa-chain idotype.
RT J. Immunol. 122:1905-1910(1979).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT
CC BIND BETA(2-I)-FRUCTOFURANOSYL MOIETIES (INULIN).
DR PIR; A01929; KWS61.
DR HSSP; P80362; 1MTL.
DR InterPro; IPR003596; Ig_MHC.
DR InterPro; IPR003596; Ig_V.

```

```

DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IgV; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 2 34 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 3 35 49 FRAMEWORK-2.
FT DOMAIN 4 50 56 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 5 57 88 FRAMEWORK-3.
FT DOMAIN 6 89 97 FRAMEWORK-4.
FT DOMAIN 7 98 107 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 23 88 FRAMEWORK-4.
FT NON_TER 108 BY SIMILARITY.
SQ SEQUENCE 108 AA; 11809 MW; FA64DA36076F2AFE CRC64;

Query Match 100.0%; Score 32; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7
DB 71 FTLTSS 77

RESULT 7
KV5R_MOUSE STANDARD; PRT; 108 AA.
ID KV5R_MOUSE
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-V region EPC 109.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=79195288; PubMed=109517;
RA Vrana M., Rudikoff S., Potter M.;
RT *The structural basis of a hapten-inhibitable kappa-chain idotype.
RT J. Immunol. 122:1905-1910(1979).
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT
CC BIND BETA(2-I)-FRUCTOFURANOSYL MOIETIES (INULIN).
DR PIR; B92808; KWS09.
DR HSSP; P01607; 1RET.
DR InterPro; IPR003596; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IgV; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 2 34 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 3 35 49 FRAMEWORK-2.
FT DOMAIN 4 50 56 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 5 57 88 FRAMEWORK-3.
FT DOMAIN 6 89 97 FRAMEWORK-4.
FT DOMAIN 7 98 107 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 23 88 FRAMEWORK-4.
FT NON_TER 108 BY SIMILARITY.
SQ SEQUENCE 108 AA; 11876 MW; 35C11BDBD0F79310 CRC64;

```

```

Query Match 100.0%; Score 32; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTSS 7
DB 71 FTLTSS 77

RESULT 8

```

KV55\_MOUSE  
ID KV55\_MOUSE STANDARD: PRT: 108 AA.  
AC P01652;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Ig kappa chain V-V region J606.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=82099361; PubMed=6798111;  
RA Johnson N., Slankard J., Paul L., Hood L.;  
RT "The complete V domain amino acid sequences of two myeloma Inulin-  
binding proteins.";  
RL J. Immunol. 128:302-307(1982).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT  
BIND BETA(2-1)-FRUCTOFURANOSYL MOIETIES (INULIN).  
DR PIR: A92811; KVAS06.  
DR HSSP: P01607; 1RE1.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003596; Ig\_V.  
DR Pfam: PF00047; 1G\_V.  
DR SMART: SM00406; 1G\_V.  
KW Immunoglobulin V region.  
FT DOMAIN 1 23  
FT DOMAIN 24 34  
FT DOMAIN 35 49  
FT DOMAIN 50 56  
FT DOMAIN 57 88  
FT DOMAIN 89 97  
FT DOMAIN 98 108  
FT DISULFID 23 88  
FT NON\_TER 108  
SQ SEQUENCE 108 AA; 11810 MW; 80E4DD31076F2AFB CRC64;

Query Match 100.0%; Score 32; DB 1; Length 108;  
Best Local Similarity 100.0%; Pred. No. 0.81;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
DB 71 FTLTSS 77

RESULT 9  
KV55\_MOUSE STANDARD: PRT: 108 AA.  
ID KV55\_MOUSE  
AC P01653;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Ig kappa chain V-V region W3082.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=82099361; PubMed=6798111;  
RA Johnson N., Slankard J., Paul L., Hood L.;  
RT "The complete V domain amino acid sequences of two myeloma Inulin-  
binding proteins.";  
RL J. Immunol. 128:302-307(1982).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT  
BIND BETA(2-1)-FRUCTOFURANOSYL MOIETIES (INULIN).  
DR PIR: B92811; KVAS82.  
DR HSSP: P80362; 1MTL.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003596; Ig\_V.  
DR InterPro: IPR003596; Ig\_V.

DR Pfam: PF00047; 1G\_V.  
DR SMART: SM00406; 1G\_V.  
KW Immunoglobulin V region.  
FT DOMAIN 1 23  
FT DOMAIN 24 34  
FT DOMAIN 35 49  
FT DOMAIN 50 56  
FT DOMAIN 57 88  
FT DOMAIN 89 97  
FT DOMAIN 98 108  
FT DISULFID 23 88  
FT NON\_TER 108  
SQ SEQUENCE 108 AA; 11850 MW; C5C145DC376F30CD CRC64;

Query Match 100.0%; Score 32; DB 1; Length 108;  
Best Local Similarity 100.0%; Pred. No. 0.81;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
DB 71 FTLTSS 77

RESULT 10  
KV1T\_HUMAN STANDARD: PRT: 109 AA.  
ID KV1T\_HUMAN  
AC P01612;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Ig kappa chain V-I region M6V.  
OS Homo sapiens (human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=83081018; PubMed=6816713;  
RA Eulitz M., Linke R.P.;  
RT "Primary structure of the variable part of an amyloidogenic  
hyperariable region of a human kappa-immunoglobulin light chain.";  
RL Hoppe-Seyler's Z. Physiol. Chem. 363:1347-1358(1982).  
CC -1- MISCELLANEOUS: ANOTHER FORM THAT LACKED RESIDUES 1-3 WAS ALSO  
FOUND.  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN.  
DR PIR: A01879; KIH0WV.  
DR HSSP: P01730; 1WIO.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003596; Ig\_V.  
DR Pfam: PF00047; 1G\_V.  
DR SMART: SM00406; 1G\_V.  
KW Immunoglobulin V region.  
FT DOMAIN 1 23  
FT DOMAIN 24 34  
FT DOMAIN 35 49  
FT DOMAIN 50 56  
FT DOMAIN 57 88  
FT DOMAIN 89 98  
FT DOMAIN 99 108  
FT DISULFID 23 88  
FT NON\_TER 109  
SQ SEQUENCE 109 AA; 11870 MW; B6ABF4515D55F5A0 CRC64;

Query Match 100.0%; Score 32; DB 1; Length 109;  
Best Local Similarity 100.0%; Pred. No. 0.82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
DB 71 FTLTSS 77

```
RESULT 11
KV3F_HUMAN STANDARD: PRT; 109 AA.
ID P01624;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ig kappa chain V-II region POM.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=76276460; PubMed=60899;
RA Klapper D.G., Capra J.D.;
RT "the amino acid sequence of the variable regions of the light chains
RL Ann. Immunol. (Paris) 127C:261-271(1976).
CC -1- MISCELLANEOUS; THIS CHAIN WAS ISOLATED FROM AN IGM WITH ANTI-GAMMA
CC GLOBULIN ACTIVITY.
DR HSSP: A01897; K3H0PM.
DR HSSP: P01789; 2MCP.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00406; IGV; 1.
KW Immunoglobulin V region.
FT DISULFID 23 89
FT NON_TER 109 109
SQ SEQUENCE 109 AA; 11922 MW; 62821DDC6A8BA86 CRC64;

Query Match 100.0%; Score 32; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 0.82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLLISS 7
Db 72 FTLLISS 78

RESULT 12
KV1A_MOUSE STANDARD: PRT; 114 AA.
ID P01632;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-I region S107A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81241357; PubMed=6788890;
RA Kwan S.-P., Rudikoff S., Seidman J.G., Leder P., Scharff M.D.;
RT "Nucleic acid and protein sequences of phosphocholine-binding light
RT chains".
RL J. Exp. Med. 153:1366-1370(1981).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL; U29423; AAC00033.1; -.
```

```
DR PIR: A01915; KYMS7A.
DR HSSP: P01789; 1MCP.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00406; IGV; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 23
FT DOMAIN 24 40
FT DOMAIN 41 55
FT DOMAIN 56 62
FT DOMAIN 63 94
FT DOMAIN 95 103
FT DOMAIN 104 113
FT DISULFID 23 94
FT NON_TER 114 114
SQ SEQUENCE 114 AA; 12717 MW; 32008EC8E9DBE7B CRC64;

Query Match 100.0%; Score 32; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLLISS 7
Db 77 FTLLISS 83

RESULT 13
KV4A_HUMAN STANDARD: PRT; 114 AA.
ID P01625;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-IV region Len.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=76004342; PubMed=50995;
RA Schneider M., Hilschmann N.;
RT "The primary structure of a monoclonic immunoglobulin-L-chain of
RT subgroup IV of the kappa type (Bence-Jones protein len)".
RL Hoppe-Seyler's Z. Physiol. Chem. 356:507-557(1975).
RN [2]
RP REVISION NO 9.
RA Salomon A.;
RL Submitted (AUG-1996) to the SWISS-PROT data bank.
CC -1- MISCELLANEOUS; THE C REGION OF THIS CHAIN HAS THE INV (?) MARKER.
CC -1- MISCELLANEOUS; THIS IS A BENCE-JONES PROTEIN.
DR PIR: A01903; K4HULN.
DR HSSP: P01789; 1MCP.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00406; IGV; 1.
KW Immunoglobulin V region; Bence-Jones protein.
FT DOMAIN 1 23
FT DOMAIN 24 40
FT DOMAIN 41 55
FT DOMAIN 56 62
FT DOMAIN 63 94
FT DOMAIN 95 101
FT DOMAIN 102 113
FT DISULFID 23 94
FT NON_TER 114 114
SQ SEQUENCE 114 AA; 12640 MW; 0647F1D17F236485 CRC64;

Query Match 100.0%; Score 32; DB 1; Length 114;
COMPLEMENTARITY-DETERMINING-1.
COMPLEMENTARITY-DETERMINING-2.
COMPLEMENTARITY-DETERMINING-3.
COMPLEMENTARITY-DETERMINING-4.
BY SIMILARITY.
FRAMEWORK-1.
FRAMEWORK-2.
FRAMEWORK-3.
FRAMEWORK-4.
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 12:58:04 ; Search time 172.49 Seconds

(without alignments)  
7.020 Million cell updates/sec

Title: US-09-712-819A-1

Perfect score: 32

Sequence: 1 FTLTIS 7

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

SPREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeop:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	32	100.0	107	4	Q96SA9	Q96SA9 homo sapien
2	32	100.0	107	11	Q9ERZ9	Q9ERZ9 mus musculu
3	32	100.0	108	4	Q9U883	Q9U883 homo sapien
4	32	100.0	108	4	Q9U177	Q9U177 homo sapien
5	32	100.0	108	4	Q9U170	Q9U170 homo sapien
6	32	100.0	109	4	Q9U185	Q9U185 homo sapien
7	30	93.8	1004	17	Q28412	Q28412 archaeoglob
8	29	90.6	116	4	Q96PF6	Q96PF6 homo sapien
9	29	90.6	171	12	Q69504	Q69504 human herpe
10	29	90.6	195	12	Q9J070	Q9J070 rice tungro
11	29	90.6	195	12	Q9J0S8	Q9J0S8 rice tungro
12	29	90.6	196	12	Q9J0T8	Q9J0T8 rice tungro
13	29	90.6	196	12	Q9J0T7	Q9J0T7 rice tungro
14	29	90.6	196	12	Q9J0T2	Q9J0T2 rice tungro
15	29	90.6	203	12	Q9J0T6	Q9J0T6 rice tungro
16	29	90.6	203	12	Q9J0T4	Q9J0T4 rice tungro

17	29	90.6	204	12	Q9J0T5	Q9J0T5 rice tungro
18	29	90.6	205	12	Q9J0S9	Q9J0S9 rice tungro
19	29	90.6	206	12	Q9J0S7	Q9J0S7 rice tungro
20	29	90.6	206	12	Q9J0S6	Q9J0S6 rice tungro
21	29	90.6	206	12	Q9J0S5	Q9J0S5 rice tungro
22	29	90.6	206	12	Q9J0S4	Q9J0S4 rice tungro
23	29	90.6	206	12	Q9J0S3	Q9J0S3 rice tungro
24	29	90.6	206	12	Q9J0S2	Q9J0S2 rice tungro
25	29	90.6	206	12	Q9J0S0	Q9J0S0 rice tungro
26	29	90.6	206	12	Q9J0S1	Q9J0S1 rice tungro
27	29	90.6	206	12	Q9J0S8	Q9J0S8 rice tungro
28	29	90.6	206	12	Q9J0S7	Q9J0S7 rice tungro
29	29	90.6	206	12	Q9J0S6	Q9J0S6 rice tungro
30	29	90.6	206	12	Q9J0S5	Q9J0S5 rice tungro
31	29	90.6	206	12	Q9J0S4	Q9J0S4 rice tungro
32	29	90.6	206	12	Q9J0S3	Q9J0S3 rice tungro
33	29	90.6	206	12	Q9J0S2	Q9J0S2 rice tungro
34	29	90.6	206	12	Q9J0S1	Q9J0S1 rice tungro
35	29	90.6	206	12	Q9J0S0	Q9J0S0 rice tungro
36	29	90.6	206	12	Q9J0S8	Q9J0S8 rice tungro
37	29	90.6	206	12	Q9J0S7	Q9J0S7 rice tungro
38	29	90.6	206	12	Q9J0S6	Q9J0S6 rice tungro
39	29	90.6	216	17	Q9J0S7	Q9J0S7 rice tungro
40	29	90.6	309	5	Q9J0S6	Q9J0S6 rice tungro
41	29	90.6	563	5	Q9J0S5	Q9J0S5 rice tungro
42	29	90.6	703	12	Q9J0S4	Q9J0S4 rice tungro
43	29	90.6	1344	12	Q9J0S3	Q9J0S3 rice tungro
44	29	90.6	1459	5	Q9J0S2	Q9J0S2 rice tungro
45	29	90.6	3471	12	Q9J0S1	Q9J0S1 rice tungro

## ALIGNMENTS

RESULT 1  
Q96SA9 ID Q96SA9 PRELIMINARY; PRT; 107 AA.  
AC Q96SA9; 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)  
DR 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE ANTI-STREPTOCOCCAL/ANTI-MYOSIN IMMUNOGLOBULIN KAPPA LIGHT CHAIN  
DE VARIABLE REGION (FRAGMENT).  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98375893; PubMed=9712075;  
RA Adderson E.E., Shikhan A.R., Ward K.E., Cunningham M.W.;  
RT \*Molecular analysis of polyclonal antibodies from  
RT rheumatic carditis: human anti-N-acetylglucosamine/anti-myosin  
RT antibody V region genes.";  
RL J. Immunol. 161:2020-2031(1998).  
DR EMBL; U96396; AAB68785.1; .  
FT NON\_TER 1  
FT NON\_TER 107  
SO SEQUENCE 107 AA; 11520 MW; 4BB4359CSB577P16 CRC64;

Query Match 100.0%; Score 32; DB 4; Length 107;  
Best Local Similarity 100.0%; Pred. No. 5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTIS 7  
DB 71 FTLTIS 77  
RESULT 2  
Q9ERZ9 PRELIMINARY; PRT; 107 AA.

```
AC 096829:
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE ANTI HUMAN TNF-ALPHA LIGHT CHAIN VARIABLE REGION (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen P., Deng J.B., Wang Z.L., Han H., Yao L.B., Su C.Z.;
RT "Cloning and sequencing of the light chain fragment of variable region
RT genes of an anti-hTNF- $\alpha$  monoclonal antibody."
RL J. Cell. Mol. Immunol. 12:21-26(1996).
RN [12]
RP SEQUENCE FROM N.A.
RA Chen P., Deng J.B., Wang Z.L., Han H., Yao L.B., Su C.Z.;
RT "Construction and sequencing of the single-chain antibody gene of a
RT human TNF- $\alpha$  specific monoclonal antibody."
RL T1 4 Chun I Ta Hsueh Hsueh Pao 19:373-376(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Chen P., Deng J.B., Wang Z.L., Han H., Yao L.B., Su C.Z.;
RT Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF262753; AAG23804.1; -.
DR HSSP: P80362; 1WTU.
DR InterPro: IPR003599; 1g.
DR InterPro: IPR003506; 1g_MHC.
DR InterPro: IPR003596; 1g_V.
DR Pfam: PF00047; 1g; 1.
DR SMART: SM00409; 1g; 1.
DR SMART: SM00406; 1g; 1.
FT NON_TER 107
SQ SEQUENCE 107 AA; 11784 MW; 2B15EEA6504A26C3 CRC64;

Query Match
Best Local Similarity 100.0%; Score 32; DB 11; Length 107;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTRISS 7
DB 74 FTLTRISS 80

RESULT 3
O9UL83 PRELIMINARY; PRT: 108 AA.
AC O9UL83:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE MYOSIN-REACTIVE IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION
DE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98277139; PubMed=9614934;
Wu X., Liu B., Van der Merwe P.L., Kallis N.N., Berney S.M.,
Young D.C.;
RA "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus."
RT Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL: AF035031; AAD56267.1; -.
DR HSSP: P80362; 1WTU.
DR InterPro: IPR003506; 1g_MHC.
DR InterPro: IPR003596; 1g_V.
DR Pfam: PF00047; 1g; 1.
DR SMART: SM00406; 1g; 1.
FT NON_TER 108
SQ SEQUENCE 108 AA; 11738 MW; C06681716C4D16F3 CRC64;
```

```
FT NON_TER 1
FT NON_TER 108
SQ SEQUENCE 108 AA; 11834 MW; 9F9C5A92BBA96EBA CRC64;

Query Match
Best Local Similarity 100.0%; Score 32; DB 4; Length 108;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTRISS 7
DB 71 FTLTRISS 77

RESULT 4
O9UL77 PRELIMINARY; PRT: 108 AA.
AC O9UL77:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE MYOSIN-REACTIVE IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION
DE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98277139; PubMed=9614934;
Wu X., Liu B., Van der Merwe P.L., Kallis N.N., Berney S.M.,
Young D.C.;
RA "Myosin-reactive autoantibodies in rheumatic carditis and normal
RT fetus."
RT Clin. Immunol. Immunopathol. 87:184-192(1998).
DR EMBL: AF035037; AAD56273.1; -.
DR HSSP: P01607; 1RET.
DR InterPro: IPR003506; 1g_MHC.
DR InterPro: IPR003596; 1g_V.
DR Pfam: PF00047; 1g; 1.
DR SMART: SM00406; 1g; 1.
FT NON_TER 108
SQ SEQUENCE 108 AA; 11738 MW; C06681716C4D16F3 CRC64;

Query Match
Best Local Similarity 100.0%; Score 32; DB 4; Length 108;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTRISS 7
DB 71 FTLTRISS 77

RESULT 5
O9UL70 PRELIMINARY; PRT: 108 AA.
AC O9UL70:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE MYOSIN-REACTIVE IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION
DE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98277139; PubMed=9614934;
Wu X., Liu B., Van der Merwe P.L., Kallis N.N., Berney S.M.,
Young D.C.;
```

```

RT      *Myosin-reactive autoantibodies in rheumatic carditis and normal
RT      fetus.":
RL      Clin. Immunol. Immunopathol. 87:184-192(1998).
DR      EMBL: AF035044; AAD56280.1; -.
DR      HSPSP; P01607; IRE1.
DR      InterPro: IPR003006; Ig_MHC.
DR      InterPro: IPR003596; Ig_V.
DR      Pfam: PF00047; Ig; 1.
DR      SMART: SM00406; IGV; 1.
FT      NON_TER 1
FT      NON_TER 108
SO      SEQUENCE 108 AA; 11633 MW; B7BEDC3E41FCCA37 CRC64;

Query Match
Best Local Similarity 100.0%; Score 32; DB 4; Length 108;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FTLTSS 7
DB      71 FTLTSS 77

RESULT  6
ID      09UL85      PRELIMINARY;      PRT;      109 AA.
AC      09UL85;
DT      01-MAY-2000 (TEMBLrel. 13, Created)
DT      01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT      01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE      MYOSIN-REACTIVE IMMUNOGLOBULIN KAPPA CHAIN VARIABLE REGION
DE      (FRAGMENT).
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=9827139; PubMed=9614934;
RA      Wu X., Liu B., Van der Merwe P.L., Kalis N.N., Berney S.M.,
RA      Young D.C.;
RT      *Myosin-reactive autoantibodies in rheumatic carditis and normal
RT      fetus.":
RL      Clin. Immunol. Immunopathol. 87:184-192(1998).
DR      EMBL: AF035029; AAD56265.1; -.
DR      HSPSP; P80362; IWTU.
DR      InterPro: IPR003006; Ig_MHC.
DR      InterPro: IPR003596; Ig_V.
DR      Pfam: PF00047; Ig; 1.
DR      SMART: SM00406; IGV; 1.
FT      NON_TER 1
FT      NON_TER 109
SO      SEQUENCE 109 AA; 11761 MW; FB1E43E7C7AFACCC CRC64;

Query Match
Best Local Similarity 100.0%; Score 32; DB 4; Length 109;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FTLTSS 7
DB      71 FTLTSS 77

RESULT  7
ID      028412      PRELIMINARY;      PRT;      1004 AA.
AC      028412;
DT      01-JAN-1998 (TEMBLrel. 05, Created)
DT      01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT      01-AUG-1998 (TEMBLrel. 07, Last annotation update)
DE      HYPOTHETICAL I15.0 KDA PROTEIN.
GN      AFI867.

```

```

OS      Archaeoglobus fulgidus.
OC      Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC      Archaeoglobus.
OX      NCBI_TaxID=2234;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      SPRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX      MEDLINE=98049343; PubMed=9389475;
RA      Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA      Retchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA      Richardson D.L., Kertlavage A.R., Graham D.E., Kyriades N.C.,
RA      Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA      Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA      Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA      Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uutterback T.,
RA      Cotton M.D., Spriggs T., Arllich P., Kaine B.P., Sykes S.M.,
RA      Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA      Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA      Venter J.C.;
RT      *The complete genome sequence of the hyperthermophilic, sulphate-
RT      reducing archaeon Archaeoglobus fulgidus.*;
RL      Nature 390:364-370(1997).
DR      EMBL: AE000974; AAB89390.1; -.
DR      TIGR: AFI867; -.
KW      Hypothetical protein; Complete proteome.
SO      SEQUENCE 1004 AA; 115036 MW; 10DB5A7A3E900F6B CRC64;

```

```

Query Match
Best Local Similarity 93.8%; Score 30; DB 17; Length 1004;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 FTLTSS 7
DB      402 FTLTSS 408

```

```

RESULT  8
ID      096PF6      PRELIMINARY;      PRT;      116 AA.
AC      096PF6;
DT      01-DEC-2001 (TEMBLrel. 19, Created)
DT      01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT      01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE      KAPPA 1 LIGHT CHAIN VARIABLE REGION (FRAGMENT).
DE      KAPPA 1 LIGHT CHAIN VARIABLE REGION (FRAGMENT).
GN      SDNK1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=21561171; PubMed=11468171;
RA      Comenzo R.L., Zhang Y., Martinez C., Osman K., Herrera G.A.;
RT      *The tropism of organ involvement in primary systemic amyloidosis:
RT      contributions of Ig V(L) germ line gene use and clonal plasma cell
RT      burden.*;
RL      Blood 98:714-720(2001).
DR      EMBL: AF361758; AAK51465.1; -.
FT      NON_TER 1
FT      NON_TER 116
SO      SEQUENCE 116 AA; 12735 MW; E796FC2217BFCF57 CRC64;

```

```

Query Match
Best Local Similarity 90.6%; Score 29; DB 4; Length 116;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 FTLTSS 7
DB      71 FTLTSS 77

```

```
RESULT 9
ID 069504 PRELIMINARY; PRT: 171 AA.
AC 069504;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-1998 (TREMBLrel. 06, Last annotation update)
DE U23 PROTEIN.
GN U23.
OS Human herpesvirus 7.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae.
OX NCBI_TaxID=10372;
RN NCBI_TaxID=10372;
RP SEQUENCE FROM N.A.
RC STRAIN=J1;
RA Nicholas J.;
RL Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RR;
RA Megaw A.G., Rapaport D., Avidor B., Frenkel N., Davison A.J.;
RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=RR;
RA Megaw A.G., Frenkel N., Davison A.J.;
RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: U43400; AAC54684.1; -.
DR EMBL: AF037218; AAC40736.1; -.
SQ SEQUENCE 171 AA; 18663 MW; ADE85FA5909D8C20 CRC64;

Query Match 90.6%; Score 29; DB 12; Length 171;
Best Local Similarity 85.7%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTIS 7
Db 98 FTLTIST 104

RESULT 10
ID 09J0T0 PRELIMINARY; PRT: 195 AA.
AC 09J0T0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE COAT PROTEIN 1 (FRAGMENT).
GN Cpl.
OS Rice tungro spherical virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Seguliviridae;
OC Waiikavirus.
OX NCBI_TaxID=35287;
RN NCBI_TaxID=35287;
RP SEQUENCE FROM N.A.
RC STRAIN=PC4111;
RA Azzam O.I., Yambao M.L.M., Muhsin M., McNally K.L., Umadhay K.M.L.;
RT "Genetic diversity of rice tungro spherical virus in tungro-endemic
RT provinces of the Philippines and Indonesia.";
RL Arch. Virol. 145:0-0(2000).
DR EMBL: AF223078; AAF65289.1; -.
FT NON_TER 1
FT NON_TER 195
SQ SEQUENCE 195 AA; 20789 MW; 0EA7DD0705357C7 CRC64;
```

```
Query Match 90.6%; Score 29; DB 12; Length 195;
Best Local Similarity 85.7%; Pred. No. 46;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTIS 7
```

```
Db 74 FTLTIS 80

RESULT 11
ID 09J0S8 PRELIMINARY; PRT: 195 AA.
AC 09J0S8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE COAT PROTEIN 1 (FRAGMENT).
GN Cpl.
OS Rice tungro spherical virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Seguliviridae;
OC Waiikavirus.
OX NCBI_TaxID=35287;
RN NCBI_TaxID=35287;
RP SEQUENCE FROM N.A.
RC STRAIN=PC8811;
RA Azzam O.I., Yambao M.L.M., Muhsin M., McNally K.L., Umadhay K.M.L.;
RT "Genetic diversity of rice tungro spherical virus in tungro-endemic
RT provinces of the Philippines and Indonesia.";
RL Arch. Virol. 145:0-0(2000).
DR EMBL: AF223080; AAF65291.1; -.
FT NON_TER 1
FT NON_TER 195
SQ SEQUENCE 195 AA; 20865 MW; 11B3D6D0704450D7 CRC64;
```

```
Query Match 90.6%; Score 29; DB 12; Length 195;
Best Local Similarity 85.7%; Pred. No. 46;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTIS 7
Db 74 FTLTIS 80
```

```
RESULT 12
ID 09J0T8 PRELIMINARY; PRT: 196 AA.
AC 09J0T8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE COAT PROTEIN 1 (FRAGMENT).
GN Cpl.
OS Rice tungro spherical virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Seguliviridae;
OC Waiikavirus.
OX NCBI_TaxID=35287;
RN NCBI_TaxID=35287;
RP SEQUENCE FROM N.A.
RC STRAIN=IB4111;
RA Azzam O.I., Yambao M.L.M., Muhsin M., McNally K.L., Umadhay K.M.L.;
RT "Genetic diversity of rice tungro spherical virus in tungro-endemic
RT provinces of the Philippines and Indonesia.";
RL Arch. Virol. 145:0-0(2000).
DR EMBL: AF223070; AAF65281.1; -.
FT NON_TER 1
FT NON_TER 196
SQ SEQUENCE 196 AA; 20917 MW; CF8299B7E3D1F6A0 CRC64;
```

```
Query Match 90.6%; Score 29; DB 12; Length 196;
Best Local Similarity 85.7%; Pred. No. 46;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTIS 7
Db 75 FTLTIS 81
```



```

RESULT 13
09J0T7 PRELIMINARY; PRT; 196 AA.
AC 09J0T7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE COAT PROTEIN 1 (FRAGMENT).
GN Cpl.
OS Rice tungro spherical virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Seguiviridae;
OC Maikavirus.
OX NCBI_TaxID=35287;
RN [1]
RP SEQUENCE FROM N.A.
RA Azzam O.I., Yambao M.L.M., Muhsin M., McNally K.L., Umadhay K.M.L.;
RT "Genetic diversity of rice tungro spherical virus in tungro-endemic
RT provinces of the Philippines and Indonesia.";
RL Arch. Virol. 145:0-0(2000).
DR EMBL; AF223071; AAF65282.1; -.
FT NON_TER 1
FT MON_TER 196
SQ SEQUENCE 196 AA; 20917 MW; CF8299B7E3D1F6A0 CRC64;

```

```

Query Match
Best Local Similarity 90.6%; Score 29; DB 12; Length 196;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 FTLTIS 7
DB 75 FTLTIS 81

```

```

RESULT 14
09J0T7 PRELIMINARY; PRT; 196 AA.
AC 09J0T7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE COAT PROTEIN 1 (FRAGMENT).
GN Cpl.
OS Rice tungro spherical virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Seguiviridae;
OC Maikavirus.
OX NCBI_TaxID=35287;
RN [1]
RP SEQUENCE FROM N.A.
RA Azzam O.I., Yambao M.L.M., Muhsin M., McNally K.L., Umadhay K.M.L.;
RT "Genetic diversity of rice tungro spherical virus in tungro-endemic
RT provinces of the Philippines and Indonesia.";
RL Arch. Virol. 145:0-0(2000).
DR EMBL; AF223076; AAF65287.1; -.
FT NON_TER 1
FT MON_TER 196
SQ SEQUENCE 196 AA; 20961 MW; 1E128B361CD0C743D CRC64;

```

```

Query Match
Best Local Similarity 90.6%; Score 29; DB 12; Length 196;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 FTLTIS 7
DB 75 FTLTIS 81

```

```

ID 09J0T6 PRELIMINARY; PRT; 203 AA.
AC 09J0T6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE COAT PROTEIN 1 (FRAGMENT).
GN Cpl.
OS Rice tungro spherical virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Seguiviridae;
OC Maikavirus.
OX NCBI_TaxID=35287;
RN [1]
RP SEQUENCE FROM N.A.
RA Azzam O.I., Yambao M.L.M., Muhsin M., McNally K.L., Umadhay K.M.L.;
RT "Genetic diversity of rice tungro spherical virus in tungro-endemic
RT provinces of the Philippines and Indonesia.";
RL Arch. Virol. 145:0-0(2000).
DR EMBL; AF223072; AAF65283.1; -.
FT NON_TER 1
FT MON_TER 203
SQ SEQUENCE 203 AA; 21647 MW; E8CC9A6A3A1E1722 CRC64;

```

```

Query Match
Best Local Similarity 90.6%; Score 29; DB 12; Length 203;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 FTLTIS 7
DB 82 FTLTIS 88

```

Search completed: July 15, 2002, 13:22:39  
Job time: 1475 sec

RESULT 15  
09J0T6



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: July 15, 2002, 12:51:04 ; Search time 228.39 Seconds  
(Without alignments)  
3.404 Million cell updates/sec

Title: US-09-712-819A-1  
Perfect score: 32  
Sequence: 1 FLVITSS 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: A\_Geneseq\_032802:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	100.0	32	17	AA87043 Human group I light
2	32	100.0	32	20	AA52745 Humanised ATR-5 L
3	32	100.0	32	22	AA898286 Anti-A33 antigen i
4	32	100.0	32	22	AA898288 Anti-A33 antigen i
5	32	100.0	32	22	AA897666 A33 antigen bindin
6	32	100.0	32	22	AA897668 A33 antigen bindin
7	32	100.0	74	15	AA862821 Human cytochrome10v1
8	32	100.0	74	19	AA862805 Amino acid sequenc
9	32	100.0	75	15	AA862823 Human cytochrome10v1
10	32	100.0	76	20	AA860981 Variable kappa lig
11	32	100.0	82	19	AA862807 Amino acid sequenc

12	32	100.0	84	17	AA814931 Monoclonal antibod
13	32	100.0	84	17	AA89878 Anti-acid sequenc
14	32	100.0	86	19	AA862806 Anti-RSV F protein
15	32	100.0	88	19	AA859619 Partial peptide fr
16	32	100.0	88	21	AA856654 Partial peptide fr
17	32	100.0	88	21	AA856655 Partial peptide fr
18	32	100.0	88	21	AA856656 Partial peptide fr
19	32	100.0	88	21	AA856657 Partial peptide fr
20	32	100.0	88	21	AA856658 Partial peptide fr
21	32	100.0	88	21	AA856659 Partial peptide fr
22	32	100.0	88	21	AA856672 Amino acid sequenc
23	32	100.0	94	19	AA862808 Light chain variab
24	32	100.0	94	22	AA867510 Human K102 VL
25	32	100.0	95	13	AA828741 Human K102 VL
26	32	100.0	95	15	AA857478 Human K102 VL
27	32	100.0	95	16	AA872061 OF7K.11 VK-1 L cha
28	32	100.0	95	16	AA872062 OF7K.19 VK-1 L cha
29	32	100.0	95	16	AA872063 OF7K.17 VK-1 L cha
30	32	100.0	95	16	AA872064 OF7K.16 VK-1 L cha
31	32	100.0	95	16	AA872065 OF7K.15 VK-1 L cha
32	32	100.0	95	16	AA872058 OF7K.3 VK-1 L cha
33	32	100.0	95	16	AA872059 OF7K.16 VK-1 L cha
34	32	100.0	95	16	AA872060 Human K102 VL regi
35	32	100.0	95	17	AA892081 Human autoantibody
36	32	100.0	98	22	AA880218 IgG antibody 2.1.1
37	32	100.0	101	20	AA834316 Humanized 369 lig
38	32	100.0	102	21	AA856683 Human anti-Rh(D) c
39	32	100.0	103	22	AA893599 Anti-HIV gp120 imm
40	32	100.0	104	17	AA854318 Anti-HIV gp120 imm
41	32	100.0	104	17	AA854318 Anti-HIV gp120 imm
42	32	100.0	104	17	AA854318 Anti-HIV gp120 imm
43	32	100.0	104	21	AA852803 Amino acid sequenc
44	32	100.0	104	21	AA852803 Anti-gp120 antibod
45	32	100.0	105	15	AA854310 Anti-HIV gp120 imm

## ALIGNMENTS

RESULT 1  
ID AA87043 standard; Peptide: 32 AA.  
XX  
AC AA87043:  
XX  
DT 25-JUN-1996 (first entry)  
XX  
DE Human group I light chain framework 3.  
XX  
KW Humanised antibody: Interleukin-5; IL-5; recombinant antibody;  
KW antibody engineering; monoclonal antibody; Mab; 39D10; CDR;  
KW complementarity determining region; light chain; framework;  
KW eosinophilia; allergy; asthma.  
XX  
OS Homo sapiens.  
XX  
PN W09535375-A1.  
XX  
PD 28-DEC-1995.  
XX  
PF 16-JUN-1995; 95MO-GB01411.  
XX  
PR 17-JUN-1994; 94GB-0012230.  
XX  
PA (CLUT ) CELUTECH THERAPEUTICS LTD.  
XX  
PI Athwal DS, Bodmer MW, Emilage JS;  
XX  
DR WPI: 1996-058412/06.  
XX  
PT Anti-human IL-5 recombinant antibody - useful for preventing or  
PT reducing eosinophilia and for treating certain allergic diseases,  
PT esp. asthma

```

XX Example 3; Fig 3; 69pp; English.
PS
XX Framework regions (AAR87041-44) of human group I (gpl) germ line
CC antibody light chain showed homology to corresponding regions
CC (AAR87045-48, respectively) of the rat anti-human interleukin-5
CC monoclonal antibody 39D10 light chain (see AAR87040). This homology
CC was utilized in the prodn. of a humanised 39D10 VL (AAR87057) in
CC which rat 39D10 VL complementarity determining regions were grafted
CC into the human gpl framework.
XX
SQ Sequence 32 AA:

Query Match 100.0%; Score 32; DB 17; Length 32;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTIS 7
Db 15 ftlitiss 21

RESULT 2
AAV52745
ID AAV52745 standard; Peptide; 32 AA.
XX
AC AAV52745;
XX
DT 26-JAN-2000 (first entry)
DE Humanised ATR-5 L chain V region FR3 for "a".
XX
KW Human tissue factor; TF; humanised; antibody; mouse monoclonal antibody;
KW ATR-2; ATR-3; ATR-4; ATR-5; ATR-7; ATR-8; thrombotic disease; DIC;
KW disseminated intravascular coagulation; immunogenicity; chimeric.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN MO9951743-A1.
XX
PD 14-OCT-1999.
XX
PF 02-APR-1999; 99WO-JP01768.
XX
PR 03-APR-1998; 98JP-0091850.
XX
PA (CHUS ) CHUGAI SEIYAKU KK.
XX
PI Sato K, Adachi H, Yabuta N;
XX
PI Sato K, Adachi H, Yabuta N;
XX
DR MPI: 1999-620204/53.
XX
PT Humanised antibody recognizing human tissue factor, used for treatment
PT of disseminated intravascular coagulation -
XX
PS Claim 17; Page 270; 291pp; Japanese.
XX
XX The present invention describes chimeric antibody (Ab) heavy (H) chains
CC containing the variable region of the H chain of a mouse monoclonal Ab
CC recognising human tissue factor (hrf) and the constant region of the H
CC chain of a human Ab. The variable region is one of six specified
CC sequences (which are the H chain variable regions from mouse monoclonal
CC Ab's ATR-2,3,4,5,7 or 8). Also described are chimeric Ab light (L)
CC chains containing the variable region of the L chain of a mouse
CC monoclonal Ab recognising human tissue factor (hrf) and the constant
CC region of the L chain of a human Ab, the variable region being one of six
CC specified sequences (which are the L chain variable regions from mouse
CC monoclonal Ab's ATR-2,3,4,5,7 or 8). The chimeric Ab's can be used for
CC the treatment and prevention of thrombotic disease, especially of
CC disseminated intravascular coagulation (DIC). The humanised antibody has
CC the high hrf binding activity of the mouse monoclonal antibody but

```

```

CC greatly reduced immunogenicity. AA233001 to AA233091 and Y527007 to
CC AAV52767 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 32 AA:

Query Match 100.0%; Score 32; DB 20; Length 32;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTIS 7
Db 15 ftlitiss 21

RESULT 3
AAB98286
ID AAB98286 standard; Peptide; 32 AA.
XX
AC AAB98286;
XX
DT 20-AUG-2001 (first entry)
DE Anti-A33 antigen immunoglobulin VL FR3 SEQ ID NO:92.
XX
KW Human; rabbit; humanised; A33 antigen; anti-A33 antigen antibody;
KW immunoglobulin; complementarity determining region; CDR; cancer;
KW cytostatic; anticancer; colon cancer; stomach cancer.
XX
OS Homo sapiens.
XX
PN MO200130393-A2.
XX
PD 03-MAY-2001.
XX
PF 20-OCT-2000; 2000WO-US29289.
XX
PR 22-OCT-1999; 99US-0425638.
XX
PR 04-APR-2000; 2000US-0543004.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PA (SLOK) SLOAN KETTERING INST CANCER RES.
XX
PA (SCRI) SCRIPPS RES INST.
XX
PI Barbas CF, Rader C, Riter G, Welt S, Old LJ;
XX
PI Barbas CF, Rader C, Riter G, Welt S, Old LJ;
XX
DR MPI: 2001-328613/34.
XX
PT Treating cancers, particularly of stomach and colon, that express A33
PT antigen by administering conjugate of anticancer agent with specific
PT immunoglobulin product -
XX
PS Claim 16; Page 40; 85pp; English.
XX
XX The present invention describes a method for treating cancers that
CC express the A33 antigen. The method comprises administering an
CC anticancer agent (I) conjugated to an immunoglobulin product (II) that
CC binds specifically to A33 and contains one or more of 13 specified
CC complementarity determining regions (CDRs), given in AAB98262 to
CC AAB98274. (I) has cytostatic activity. The method can be used for
CC treating colon and stomach cancers. (II), or the nucleic acid encoding
CC it, can be used directly, in unconjugated form, for immunotherapy of
CC cancer, and, when labeled, for detection or diagnosis of diseases
CC associated with A33 expression. AA42218 to AA82254 and AAB98230 to
CC AAB98321 represent sequences used in the exemplification of the
CC present invention.
XX
SQ Sequence 32 AA:

Query Match 100.0%; Score 32; DB 22; Length 32;
Best Local Similarity 100.0%; Pred. No. 3.2;

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTIS 7  
 |||||  
 Db 15 ftlctss 21

## RESULT 4

AA98288  
 ID AAB98288 standard; Peptide: 32 AA.  
 AC AAB98288;  
 XX  
 XX 20-AUG-2001 (first entry)  
 DE Anti-A33 antigen immunoglobulin VL FR3 SEQ ID NO:94.  
 DE  
 XX Human; rabbit; humanised; A33 antigen; anti-A33 antigen antibody;  
 KW immunoglobulin; complementarity determining region; CDR; cancer;  
 KW cytosolic; anticancer; colon cancer; stomach cancer.  
 OS Homo sapiens.  
 XX  
 XX WO200130393-A2.  
 XX 03-MAY-2001.  
 XX 20-OCT-2000; 2000MO-US29289.  
 XX  
 XX 22-OCT-1999; 99US-0425638.  
 XX 04-APR-2000; 2000US-0543004.  
 XX  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 PA (SLOC) SLOAN KETTERING INST CANCER RES.  
 PA (SCRI) SCRIPPS RES INST.  
 XX  
 PI Barbas CF, Rader C, Rittter G, Welt S, Old LJ;  
 DR WPI: 2001-328613/34.  
 XX  
 PT Treating cancers, particularly of stomach and colon, that express A33  
 PT antigen by administering conjugate of anticancer agent with specific  
 PT immunoglobulin product -  
 XX  
 XX Claim 16; Page 40; 85pp; English.  
 XX  
 CC The present invention describes a method for treating cancers that  
 CC express the A33 antigen. The method comprises administering an  
 CC anticancer agent (I) conjugated to an immunoglobulin product (II) that  
 CC binds specifically to A33 and contains one or more of 13 specified  
 CC complementarity determining regions (CDRs), given in AAB98262 to  
 CC AAB98274. (I) has cytosolic activity. The method can be used for  
 CC treating colon and stomach cancers. (II), or the nucleic acid encoding  
 CC it, can be used directly, in unconjugated form, for immunotherapy of  
 CC cancer, and, when labeled, for detection or diagnosis of diseases  
 CC associated with A33 expression. AAH22218 to AAH22254 and AAB98230 to  
 CC AAB98321 represent sequences used in the exemplification of the  
 CC present invention.  
 XX  
 XX Sequence 32 AA;

Query Match 100.0%; Score 32; DB 22; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 3.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLTIS 7  
 |||||  
 Db 15 ftlctss 21

RESULT 5  
 AAB97666

ID AAB97666 standard; Peptide: 32 AA.

AC AAB97666;

DT 08-AUG-2001 (first entry)

DE A33 antigen binding immunoglobulin product VLFR3 peptide SEQ ID NO:92.

KW Chimeric antibody; humanised; humanisation; mammalian antibody; antigen;  
 KW immunoreact; anti A33 antigen antibody; immunoglobulin.

OS Homo sapiens.

PN WO200131065-A1.

PD 03-MAY-2001.

PF 20-OCT-2000; 2000MO-US29026.

PR 22-OCT-1999; 99US-0425638.

PR 04-APR-2000; 2000US-0543004.

PA (SCRI) SCRIPPS RES INST.

PI Barbas CF, Rader C;

DR WPI: 2001-328657/34.

PT Preparing humanized rabbit antibodies that specifically immunoreact  
 PT with a particular antigen using display technology for expressing  
 PT libraries of antibody domains and fine tuning variable domain regions -

PS Example 9; Page 39; 62pp; English.

CC The present invention describes a method for preparing a humanised rabbit  
 CC antibody that specifically immunoreacts with a particular antigen. The  
 CC method comprises expressing a library of antibodies comprising one or  
 CC more complementarity determining region (CDR) from the variable domain  
 CC sequences that specifically immunoreact with the antigen grafted into  
 CC framework regions from humans, and selecting the antibodies that react  
 CC with the antigen. The method is useful for humanising non-human  
 CC mammalian antibodies, which can be used for the treatment of a variety  
 CC of diseases. The present sequence represents an A33 antigen binding  
 CC immunoglobulin product VLFR3 peptide which is given in an example from  
 CC the present invention.

XX Sequence 32 AA;

QY 1 FTLTIS 7  
 |||||  
 Db 15 ftlctss 21

Query Match 100.0%; Score 32; DB 22; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 3.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 6  
 AAB97668  
 ID AAB97668 standard; Peptide: 32 AA.  
 AC AAB97668;  
 XX  
 XX 08-AUG-2001 (first entry)

DE A33 antigen binding immunoglobulin product VLFR3 peptide SEQ ID NO:94.

KW Chimeric antibody; humanised; humanisation; mammalian antibody; antigen;  
 KW immunoreact; anti A33 antigen antibody; immunoglobulin.

OS Homo sapiens.

PN WO20011065-A1.  
 XX  
 PD 03-MAY-2001.  
 XX  
 PP 20-OCT-2000; 2000WO-US29026.  
 XX  
 PR 22-OCT-1999; 99US-0425638.  
 PR 04-APR-2000; 2000US-0543004.  
 XX  
 PA (SCRI ) SCRIPPS RES INST.  
 XX  
 PI Barbas CF, Rader C;  
 DR WPI: 2001-328657/34.  
 XX  
 XX Preparing humanized rabbit antibodies that specifically immunoreact  
 PT with a particular antigen using display technology for expressing  
 PT libraries of antibody domains and fine tuning variable domain regions -  
 XX  
 PS Example 9; Page 39; 62pp; English.  
 XX  
 CC The present invention describes a method for preparing a humanised rabbit  
 CC antibody that specifically immunoreacts with a particular antigen. The  
 CC method comprises expressing a library of antibodies comprising one or  
 CC more complementarily determining region (CDR) from the variable domain  
 CC sequences that specifically immunoreact with the antigen grafted into  
 CC framework regions from humans, and selecting the antibodies that react  
 CC with the antigen. The method is useful for humanising non-human  
 CC mammalian antibodies, which can be used for the treatment of a variety  
 CC of diseases. The present sequence represents an A33 antigen binding  
 CC immunoglobulin product VLER3 peptide which is given in an example from  
 CC the present invention.  
 XX  
 SQ Sequence 32 AA:  
 Query Match 100.0%; Score 32; DB 22; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 3.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FTLTIS 7  
 DB 15 flltiss 21  
 RESULT 7  
 AAR62921  
 ID AAR62921 standard; protein: 74 AA.  
 XX  
 AC AAR62921;  
 XX  
 DT 18-JUL-1995 (first entry)  
 XX  
 DE Human cytomagalovirus antibody light chain variable region.  
 XX  
 KW Human cytomagalovirus: antibody light chain variable region;  
 KW HMCV; Immunocassay; Immunotherapy.  
 OS Homo sapiens.  
 XX  
 PN WO9425490-A.  
 PD 10-NOV-1994.  
 XX  
 PP 29-APR-1994; 94WO-US04705.  
 XX  
 PR 30-APR-1993; 93US-0055985.  
 XX  
 PA (SCRI ) SCRIPPS RES INST.  
 XX  
 PI Barbas C, Burton R, Burton DR, Williamson A, Burton R;  
 DR WPI: 1994-358194/44.

XX  
 PT Human monoclonal antibodies (Mabs) against human cytomagalovirus  
 PT - also nucleic acids and cell lines producing the Mabs, useful in  
 PT diagnosis and immunotherapy  
 XX  
 PS Claim 6; Page 150; 171pp; English.  
 XX  
 CC AAR62618-R62920 and AAR62921-R62923 are human cytomagalovirus (HMCV)  
 CC antibody heavy chain and light chain variable regions respectively.  
 CC A monoclonal antibody (Mab) containing a pair of these regions can be  
 CC used to detect HCV, and anti-human HCVy Abs in human patients via  
 CC a competitive immunoassay. The Mabs may also be useful in  
 CC immunotherapy.  
 XX  
 SQ Sequence 74 AA:  
 Query Match 100.0%; Score 32; DB 15; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 8.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FTLTIS 7  
 DB 34 flltiss 40  
 RESULT 8  
 AAM62805  
 ID AAM62805 standard; Peptide: 74 AA.  
 XX  
 AC AAM62805;  
 XX  
 DT 23-SEP-1998 (first entry)  
 XX  
 DE Amino acid sequence of a human antibody fragment.  
 XX  
 KW Human: immunoglobulin; Ig; transgenic; non-human mammal;  
 KW inactivated endogenous Ig locus; B-cell development;  
 KW human heavy chain Ig locus; micro constant region; J-H; D-H; V-H gene;  
 KW kappa light chain Ig locus; kappa constant region; J-kappa gene; V-kappa;  
 KW production; antibody.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9824893-A2.  
 PD 11-JUN-1998.  
 XX  
 PP 03-DEC-1997; 97WO-US23091.  
 PR 03-DEC-1996; 96US-0759620.  
 XX  
 PA (ABGE-) ABGENIX INC.  
 XX  
 PI Green L, Jakobovits A, Klapholz S, Kucherlapati R;  
 PI Mendez M;  
 DR WPI: 1998-333314/29.  
 XX  
 PT New transgenic non-human mammals - having an inactivated  
 PT immunoglobulin locus and a near complete human immunoglobulin locus,  
 PT used for production of human antibodies  
 XX  
 PS Disclosure; Page 77; 128pp; English.  
 XX  
 CC AAM62793-822 represent fragments of human antibodies produced by  
 CC transgenic Xenomice, created using the method of the invention. The  
 CC specification describes a transgenic non-human mammal which has genome  
 CC modifications that comprise an inactivated endogenous immunoglobulin (Ig)  
 CC locus, so that the mammal does not display normal B-cell development. The  
 CC modified genome also has an inserted human heavy chain Ig locus in  
 CC germline configuration, the human heavy chain Ig locus comprising a human  
 CC micro constant region and regulatory and switch sequences, human J-H

genes, human D-H genes, and human V-H genes and an inserted human kappa light chain Ig locus in germline configuration, the human kappa light chain Ig locus comprising a human kappa constant region, J-kappa genes, and V-kappa genes, where the number of V-H and V-kappa genes inserted are selected to restore normal B-cell development in the mammal. The transgenic animals have a near complete human Ig locus, including both a human heavy chain locus and a human kappa light chain locus. They can be used for the production of human antibodies when exposed to particular antigens e.g. when exposed to human IL-8, EGFR or TNF- alpha the mice will produce antibodies to IL-8, EGFR or TNF- alpha respectively.

Sequence 74 AA;

Query Match 100.0%; Score 32; DB 19; Length 74;  
Best Local Similarity 100.0%; Pred. No. 8.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
| | | | |  
Db 50 flltiss 56

# RESULT 9

AA62923  
ID AAR62923 standard; protein; 75 AA.

AC AAR62923;

DT 18-JUL-1995 (first entry)

DE Human cytomagalovirus antibody light chain variable region.

KW Human cytomagalovirus; antibody light chain variable region;

KV HMCV; Immunassay; Immunotherapy.

XX Homo sapiens.

OS WO9425490-A.

PN 10-NOV-1994.

PD -29-APR-1994; 94WO-US04705.

PF 30-APR-1993; 93US-0055985.

PR (SCRI ) SCRIPPS RES INST.

PA Barbab C, Burton R, Burton DR, Williamson A, Burloni R;

PI WPI; 1994-358194/44.

DR Human monoclonal antibodies (MAbs) against human cytomagalovirus

PT - also nucleic acids and cell lines producing the MAbs, useful in

PT diagnosis and immunotherapy

PS Claim 6; Page 151; 171pp; English.

XX AAR62818-R62920 and AAR62921-R6293 are human cytomagalovirus (HMCV)

CC antibody heavy chain and light chain variable regions respectively.

CC A monoclonal antibody (Mab) containing a pair of these regions can be

CC used to detect HCMV, and anti-human HCMV Abs in human patients via

CC a competitive immunassay. The MAbs may also be useful in

CC immunotherapy.

Sequence 75 AA;

Query Match 100.0%; Score 32; DB 15; Length 75;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLTSS 7  
| | | | |  
Db 37 flltiss 43

# RESULT 10

AA60981  
ID AAM80981 standard; Protein; 76 AA.

AC AAM80981;

DT 30-MAR-1999 (first entry)

DE Variable kappa light region 012 encoded amino acid.

KW Human; epidermal growth factor receptor; tumour; EGF;

KV Transforming growth factor alpha; TGF-alpha.

XX Homo sapiens.

OS WO9850433-A2.

PN 12-NOV-1998.

PD 05-MAY-1998; 98WO-US09160.

PR 05-MAY-1997; 97US-0851362.

PA (ABGE-) ABGENIX INC.

PI Gallo M, Jakobovits A, Jia X, Yang X;

DT WPI; 1999-034712/03.

DE Humanised antibodies against epidermal growth factor receptor, EGF-r

PT - useful to treat solid tumours whilst inducing reduced immunogenic

PT or allergic effects compared to mouse or mouse-derived antibodies

PS Example 3; Page 105; 143pp; English.

XX The variable kappa light region 012 encoded amino acid was used in the

CC production of anti-epidermal growth factor receptor (EGF-r)-antibodies.

CC The antibodies can be administered therapeutically to patients (human or

CC veterinary) to treat solid tumours. EGF-r is overexpressed on many human

CC solid tumour types, and the fully human antibodies (i.e. comprising and

CC inhibit both epidermal growth factor (EGF) and transforming growth factor

CC alpha (TGF-alpha) binding to EGF-r (known to lead to cellular

CC proliferation and tumour growth). They can prevent tumour cell growth

CC and, in combination with an antineoplastic agent, may eradicate

CC established tumours. The fully human antibodies can minimise the

CC immunogenic and allergic responses intrinsic to previous mouse/rat or

CC mouse/rat-derived antibodies.

Sequence 76 AA;

Query Match 100.0%; Score 32; DB 20; Length 76;  
Best Local Similarity 100.0%; Pred. No. 8.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

# RESULT 11

AA62807  
ID AAW62807 standard; Reptide; 82 AA.

AC AAW62807;

DT 23-SEP-1998 (first entry)

DE Amino acid sequence of a human antibody fragment.  
 XX  
 XX Human; Immunoglobulin; Ig; transgenic; non-human mammal;  
 KW Inactivated endogenous Ig locus; B-cell development;  
 KW human heavy chain Ig locus; micro constant region; J-H; D-H; V-H gene;  
 KW kappa light chain Ig locus; kappa constant region; J-kappa gene; V-kappa;  
 KW production; antibody.  
 XX  
 OS Homo sapiens.  
 XX  
 XX W09824893-A2.  
 PM  
 PD 11-JUN-1998.  
 XX  
 XX 03-DEC-1997; 97MO-US23091.  
 PF  
 XX 03-DEC-1996; 96US-0759620.  
 PR  
 XX (ABGE-) ABGENIX INC.  
 PA  
 XX Green L, Jakobovits A, Klapholz S, Kucherlapati R;  
 PI Mendez M;  
 XX  
 XX WPI: 1998-333314/29.  
 DR  
 XX  
 XX New transgenic non-human mammals - having an inactivated  
 PT immunoglobulin locus and a near complete human immunoglobulin locus,  
 PT used for production of human antibodies  
 PS  
 XX Disclosure; Page 78; 128pp; English.  
 PS  
 XX AAW62793-822 represent fragments of human antibodies produced by  
 CC transgenic xenotic, created using the method of the invention. The  
 CC specification describes a transgenic non-human mammal which has genome  
 CC modifications that comprise an inactivated endogenous immunoglobulin (Ig)  
 CC locus, so that the mammal does not display normal B-cell development. The  
 CC modified genome also has an inserted human heavy chain Ig locus in  
 CC germ-line configuration, the human heavy chain Ig locus comprising a human  
 CC micro constant region and regulatory and switch sequences, human J-H  
 CC genes, human D-H genes, and human V-H genes and an inserted human kappa  
 CC light chain Ig locus in germ-line configuration, the human kappa light  
 CC chain Ig locus comprising a human kappa constant region, J-kappa genes,  
 CC and V-kappa genes, where the number of V-H and V-kappa genes inserted  
 CC are selected to restore normal B-cell development in the mammal. The  
 CC transgenic animals have a near complete human Ig locus, including both a  
 CC human heavy chain locus and a human kappa light chain locus. They can  
 CC be used for the production of human antibodies when exposed to  
 CC particular antigens e.g. when exposed to human IL-8, EGF or TNF- alpha  
 CC the mice will produce antibodies to IL-8, EGF or TNF- alpha  
 CC respectively.  
 CC  
 CC Sequence 82 AA;  
 SQ  
 Query Match 100.0%; Score 32; DB 19; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 9;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 FTLTISS 7  
 DB 58 flltiss 64  
 OY  
 DB 58 flltiss 64  
 RESULT 12  
 AAW14491  
 ID AAW14491 standard; Protein; 84 AA.  
 XX  
 XX AAW14491;  
 AC  
 XX 28-JAN-1997 (first entry)  
 DT  
 XX Monoclonal antibody D VK.  
 DE  
 XX

KW heavy; light chain; monoclonal antibody; antigen 3; marker; melanoma;  
 KW permanent human tumour cell line; tumour-associated antigen; epitope;  
 KW gastrointestinal tumour; pancreatic carcinoma; diagnostic; therapeutic;  
 KW antigen 11; Vibrio cholera; neuraminidase-resistant; ganglioside GD2.  
 XX  
 OS Synthetic.  
 OS  
 XX EP727436-A1.  
 PN  
 XX  
 XX 21-AUG-1996.  
 PD  
 XX  
 XX 21-MAR-1990; 90EP-0105322.  
 PF  
 XX 24-MAR-1989; 89DE-3909799.  
 PR  
 XX (BEHW ) BEHRINGWERKE AG.  
 PA  
 XX  
 XX Auerbach B, Bosslet K, Sedlacek H, Seemann G;  
 PI WPI: 1996-372836/38.  
 DR  
 XX N-PSDB; AAT63508.  
 DR  
 XX Monoclonal antibody to tumour-associated antigen - useful as  
 PT gastrointestinal tumour marker  
 PT  
 XX  
 XX Disclosure; Page 14; 19pp; German.  
 PS  
 XX  
 XX AAW14490-91 are the heavy and light chains (respectively) of monoclonal  
 CC antibody (MAB) D. MAB D recognises Vibrio cholera  
 CC neuraminidase-resistant epitope of ganglioside GD2, from a human melanoma  
 CC cell line. MABs A, B and C (see AAW14484-89) are mentioned in the  
 CC specification, but are not part of the claims. MABs A and B recognise  
 CC antigens 3 and 11 resp., of a permanent human tumour cell line. MAB C  
 CC also recognises an epitope of a tumour-associated antigen. These antigens  
 CC occur at high concns. in the serum of patients with gastrointestinal  
 CC tumours, e.g. pancreatic carcinoma, and are thus useful as tumour markers  
 CC for diagnostic or therapeutic purposes.  
 CC  
 CC Sequence 84 AA;  
 SQ  
 Query Match 100.0%; Score 32; DB 17; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 9.3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 FTLTISS 7  
 DB 49 flltiss 55  
 OY  
 DB 49 flltiss 55  
 RESULT 13  
 AAR9878  
 ID AAR9878 standard; Protein; 84 AA.  
 XX  
 XX AAR9878;  
 AC  
 XX 28-JAN-1997 (first entry)  
 DT  
 XX Monoclonal antibody D VK.  
 DE  
 XX Monoclonal antibody; MAB; epitope; tumour-associated antigen;  
 KW marker; antigen.  
 KW  
 XX Synthetic.  
 OS  
 XX EP727435-A1.  
 PN  
 XX 21-AUG-1996.  
 PD  
 XX 21-MAR-1990; 90EP-0105322.  
 PF  
 XX 24-MAR-1989; 89DE-3909799.  
 PR  
 XX



PA (BEHW ) BEHRINGWERKE AG.  
 XX Auerbach B, Bosslet K, Sedlacek H, Seemann G:  
 XX WPI; 1996-372835/38.  
 DR N-PSDB; AAT36666.  
 XX Monoclonal antibody to tumour-associated antigen - useful as  
 PT gastrointestinal tumour marker  
 XX Disclosure; Page 14; 199p; German.  
 XX Mab C (AAT36659-T36660) is a monoclonal antibody that recognises an  
 CC epitope of a tumour-associated antigen occurring at high concn. in  
 CC the serum of patients with gastrointestinal tumours, e.g. pancreatic  
 CC carcinoma, and is thus useful as a tumour marker for diagnostic or  
 CC therapeutic purposes.  
 CC Mabs A, B and D are mentioned in the specification, but are not  
 CC part of the claims.  
 CC Mab A (AAT36661-T36662) recognises antigen 3 of permanent human  
 CC tumour cell line.  
 CC Mab B (AAT36663-T36664) recognises antigen 11 of permanent human  
 CC tumour cell line.  
 CC Mab D (AAT36665-T36666) recognises a Vibrio cholera neuraminidase-  
 CC resistant epitope of ganglioside GD2, from a human melanoma cell  
 CC line.  
 XX Sequence 84 AA:  
 SQ  
 Query Match 100.0%; Score 32; DB 17; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 9.3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FTITISS 7  
 DB 49 fTITISS 55  
 RESULT 14  
 ID AAM62806 standard; Peptide; 86 AA.  
 XX AAM62806;  
 XX 23-SEP-1998 (first entry)  
 DE Amino acid sequence of a human antibody fragment.  
 XX Human; immunoglobulin; Ig; transgenic; non-human mammal;  
 KW inactivated endogenous Ig locus; B-cell development;  
 KW human heavy chain Ig locus; micro constant region; J-H; D-H; V-H gene;  
 KW kappa light chain Ig locus; kappa constant region; J-kappa gene; V-kappa;  
 production; antibody.  
 XX Homo sapiens.  
 XX OS  
 XX W09824893-A2.  
 XX 11-JUN-1998.  
 PD 03-DEC-1997; 97WO-US23091.  
 PF 03-DEC-1996; 96US-0759620.  
 XX 03-DEC-1996; 96US-0759620.  
 XX (ABGE-) ABGENIX INC.  
 XX Green L, Jakobovits A, Klapholz S, Kucheralapati R;  
 PI Mendez W;  
 XX WPI; 1998-333314/29.  
 DR New transgenic non-human mammals - having an inactivated

PT immunoglobulin locus and a near complete human immunoglobulin locus,  
 PT used for production of human antibodies  
 XX Disclosure; Page 78; 128pp; English.  
 XX AAM62793-822 represent fragments of human antibodies produced by  
 CC transgenic Xenomice, created using the method of the invention. The  
 CC specification describes a transgenic non-human mammal which has genome  
 CC modifications that comprise an inactivated endogenous immunoglobulin (Ig)  
 CC locus, so that the mammal does not display normal B-cell development. The  
 CC modified genome also has an inserted human heavy chain Ig locus in  
 CC germline configuration, the human heavy chain Ig locus comprising a human  
 CC micro constant region and regulatory and switch sequences, human J-H  
 CC genes, human D-H genes, and human V-H genes and an inserted human kappa  
 CC light chain Ig locus in germline configuration, the human kappa light  
 CC chain Ig locus comprising a human kappa constant region, J-kappa genes,  
 CC and V-kappa genes, where the number of V-H and V-kappa genes inserted  
 CC are selected to restore normal B-cell development in the mammal. The  
 CC transgenic animals have a near complete human Ig locus, including both a  
 CC human heavy chain locus and a human kappa light chain locus. They can  
 CC be used for the production of human antibodies when exposed to  
 CC particular antigens e.g. when exposed to human IL-8, EGFR or TNF- alpha  
 CC respectively.  
 XX Sequence 86 AA:  
 SQ  
 Query Match 100.0%; Score 32; DB 19; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 9.5;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FTITISS 7  
 DB 50 fTITISS 56  
 RESULT 15  
 ID AAM59619 standard; Protein; 88 AA.  
 XX AAM59619;  
 XX 12-OCT-1998 (first entry)  
 DE Anti-RSV F protein Hn19 light chain GL DpK9.  
 XX Monoclonal antibody; human; Hn19; engineered antibody; RSV;  
 KW respiratory syncytial virus; fusion protein; infection;  
 KW complementarity determining region; CDR; therapy; diagnosis.  
 XX Homo sapiens.  
 XX OS  
 XX FH Key Location/Qualifiers  
 FT Region 24..34  
 FT /label= CDR1  
 FT Region 50..57  
 FT /label= CDR2  
 XX W09819704-A1.  
 XX 14-MAY-1998.  
 PD 23-OCT-1997; 97WO-US19203.  
 PF 01-NOV-1996; 96US-0030149.  
 XX 01-NOV-1996; 96US-0030149.  
 XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX Deen KC, Dillon SB, Porter TG, Sweet RW;  
 XX WPI; 1998-286600/25.

PT Monoclonal antibodies reactive with Respiratory Syncytial Virus -  
 useful for detection, prevention and treatment of RSV infections

XX  
 PS Disclosure; Fig 3: 109pp; English.

CC This is the heavy chain amino acid sequence of germline (GL)  
 CC DPK9 antibody. The invention relates to the construction and use  
 CC of human MAb specific for the fusion (F) protein of respiratory  
 CC syncytial virus (RSV) to passively treat, prevent or detect RSV  
 CC infection. Hu19A, Hu19B, Hu19C and Hu19D MAb are claimed. These  
 CC are reshaped human antibodies comprising a heavy chain selected  
 CC from 19A, 19B, 19C or 19D (see AAW59615-18), which are based on the  
 CC GL DPK8 sequence, and a light chain selected from 19A, 19B, 19C or  
 CC 19D (see AAW59620-21). Such engineered antibodies are neutralising;  
 CC they inhibit virus growth in vitro and in vivo in animal models of  
 CC RSV infection. Nucleic acids encoding the human MAb recombinant  
 CC plasmids (see AAW1427-33) and host cells (e.g. COS, CHO, myeloma)  
 CC are provided.

XX  
 SQ Sequence 88 AA;

Query Match 100.0%; Score 32; DB 19; Length 88;

Best Local Similarity 100.0%; Pred. No. 9.8; Mismatches 0; Gaps 0;

Matches 7; Conservative 0; Indels 0;

OY 1 FTLRIS 7

Db 71 ftltis 77

Search completed: July 15, 2002, 12:57:54  
 Job time: 410 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 15, 2002, 13:00:57 ; Search time 95.45 Seconds  
(without alignments)  
7.047 Million cell updates/sec

Title: US-09-712-819a-2  
Perfect score: 33  
Sequence: 1 TRIFSXL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	97.0	529	2 T03879	hypothetical prote
2	30	90.9	499	2 T03877	hypothetical prote
3	30	90.9	1106	2 C87128	probable transcrip
4	30	90.9	1227	2 T50394	DNA-directed RNA p
5	29	87.9	244	2 B87417	hypothetical prote
6	29	87.9	291	2 D96998	phosphatidylserine
7	29	87.9	352	2 T32314	hypothetical prote
8	29	87.9	590	2 E90089	hypothetical prote
9	29	87.9	592	2 S42220	transcription init
10	29	87.9	1071	2 T18597	hypothetical prote
11	28	84.8	39	2 H81081	hypothetical prote
12	28	84.8	179	2 B75636	transcription regu
13	28	84.8	187	2 G70196	hypothetical prote
14	28	84.8	193	2 G84869	hypothetical prote
15	28	84.8	328	2 E88930	protein R11G11.11
16	28	84.8	381	2 T39187	neuropeptide V/pep
17	28	84.8	422	2 T21820	hypothetical prote
18	28	84.8	483	2 S25070	xylokinase (EC 2
19	28	84.8	586	2 T04716	hypothetical prote
20	28	84.8	744	2 T00652	hypothetical prote
21	28	84.8	834	2 T00512	serine/threonine-s
22	28	84.8	852	2 A00050	probable phosphoen
23	28	84.8	1317	2 A54831	nuclear pore compl
24	27	81.8	156	2 C09330	conserved hypotnet
25	27	81.8	204	2 D11648	hypothetical prote
26	27	81.8	237	2 S73263	hypothetical prote
27	27	81.8	342	2 T48045	hypothetical prote
28	27	81.8	343	2 A41748	lumican precursor
29	27	81.8	352	2 S70972	bipe protein - Esc

30	27	81.8	355	2 T10722	anthocyanidin synt
31	27	81.8	392	2 C96766	hypothetical prote
32	27	81.8	398	2 F85056	hypothetical prote
33	27	81.8	405	2 D96709	probable B-box zin
34	27	81.8	416	2 E86384	probable zinc fing
35	27	81.8	440	2 AF2934	uracil transport p
36	27	81.8	463	2 S52754	aromatic amino aci
37	27	81.8	502	2 A98348	probable purine pe
38	27	81.8	522	2 T18583	glucosylceramidase
39	27	81.8	582	2 B82805	prolyl-L-lysine synth
40	27	81.8	584	1 C8H0A	complement C8 alph
41	27	81.8	611	2 A83926	hypothetical prote
42	27	81.8	612	2 H96494	protein F7P22.2 [1
43	26	78.8	67	2 AB2780	hypothetical prote
44	26	78.8	85	2 T28292	ORF MSV131 hypote
45	26	78.8	121	2 T50205	hypothetical prote

#### ALIGNMENTS

RESULT 1

T03879 hypothetical protein F07G11.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 26-May-2000

C:Accession: T03879

R:Sammons, L.; Mohlmann, P.; Sansone, J.

submitted to the EMBL Data Library, July 1997

A:Description: The sequence of C. elegans cosmid F07G11.

A:Reference number: Z15127

A:Accession: T03879

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-529 <S>AM>

A:Cross-references: EMBL:AF016419; NID:g2291159; PIDN:AA865295.1; PID:g2291160

C:Genetics:

A:Map position: V

A:Introns: 73/2; 120/2; 155/3; 193/2; 219/3; 269/1; 392/3

A:Note: F07G11.3

C:superfamily: Caenorhabditis elegans hypothetical protein F07G11.3

Query Match

Best local similarity 97.0%; Score 32; DB 2; Length 529;

Matches 6; Conservative 1; Mismatches 0; Indels 0;

Gaps 0;

QY 1 TRIFSXL 7

DB 440 TRIFSXL 446

RESULT 2

T03877 hypothetical protein F07G11.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 26-May-2000

C:Accession: T03877

R:Sammons, L.; Mohlmann, P.; Sansone, J.

submitted to the EMBL Data Library, July 1997

A:Description: The sequence of C. elegans cosmid F07G11.

A:Reference number: Z15127

A:Accession: T03877

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-449 <S>AM>

A:Cross-references: EMBL:AF016419; NID:g2291159; PIDN:AA865296.1; PID:g2291163

C:Genetics:

A:Map position: V

A:Introns: 56/2; 103/2; 138/3; 176/2; 202/3; 246/3; 318/3; 363/3

A:Note: F07G11.4

C:superfamily: Caenorhabditis elegans hypothetical protein F07G11.3

Query Match 90.9%; Score 30; DB 2; Length 499;  
 Best Local Similarity 85.7%; Pred. No. 45;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSKL 7  
 ||:||||  
 Db 410 TRMFSL 416

# RESULT 3

C87128  
 probable transcription regulator [Imported] - Mycobacterium leprae  
 C:Species: Mycobacterium leprae  
 C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C:Accession: C87128

R:Coile, S.T.; Elgmelter, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; HC  
 R.; Davies, R.M.; Devlin, K.; Dutfoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
 cam, M.A.; Rutherford, K.M.  
 Nature 409, 1007-1011, 2001  
 A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sg  
 A:Title: Massive gene decay in the leprosy bacillus.  
 A:Reference number: A86909; MUID:21128732; PMID:11234002  
 A:Accession: C87128  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 11106 <STO>  
 A:Cross-references: GB:AL450380; NID:g13093492; PIDN:CAC30706.1; GSPDB:GN00147  
 A:Gene: ML1753

Query Match 90.9%; Score 30; DB 2; Length 1106;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSKL 7  
 ||:||||  
 Db 1084 TRIFSKL 1090

# RESULT 4

T50394  
 DNA-directed RNA polymerase subunit [Imported] - fission yeast (Schizosaccharomyces pombe  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
 C:Accession: T50394

R:Beck, A.; Borzym, K.; Reinhardt, R.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, January 1999  
 A:Reference number: 225067  
 A:Accession: T50394  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 11227 <BEC>  
 A:Cross-references: EMBL:AL136535; PIDN:CAB66435.1; GSPDB:GN00067; SPDB:SPBP23A10.07  
 C:Experimental source: strain 972h(-); clone pl p23A10  
 C:Genetics:  
 A:Gene: rpa2; SPDB:SPBP23A10.07  
 A:Map position: 2  
 C:Superfamily: DNA-directed RNA polymerase 132k polypeptide

Query Match 90.9%; Score 30; DB 2; Length 1227;  
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSKL 7  
 ||:||||  
 Db 476 TRIFSKL 482

# RESULT 5

B87417

hypothetical protein CC1353 [Imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus  
 C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C:Accession: B87417  
 R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,  
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwin, M.L.; Haft, D.H.; Ko  
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A:Title: Complete Genome Sequence of Caulobacter crescentus.  
 A:Reference number: AB7249; MUID:21173698; PMID:11259647  
 A:Accession: B87417  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-244 <STO>  
 A:Cross-references: GB:AE005673; NID:g14322702; PIDN:AAK23334.1; GSPDB:GN00148  
 C:Genetics:  
 A:Gene: CC1353

Query Match 87.9%; Score 29; DB 2; Length 244;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSK 6  
 ||:||||  
 Db 69 TRIFSK 74

# RESULT 6

D96998  
 phosphatidylserine decarboxylase [Imported] - Clostridium acetobutylicum  
 C:Species: Clostridium acetobutylicum  
 C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
 C:Accession: D96998

R:Noelling, J.; Breton, G.; Omeichenko, M.V.; Markova, K.S.; Zeng, Q.; Gibson, R.; L  
 ; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
 A:Reference number: A96900; MUID:21359325; PMID:21359325  
 A:Accession: D96998  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-291 <KUD>

A:Cross-references: GB:AF001437; PIDN:AAK78775.1; PID:g150233667; GSPDB:GN00168  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC0799  
 C:Superfamily: Escherichia coli phosphatidylserine decarboxylase

Query Match 87.9%; Score 29; DB 2; Length 291;  
 Best Local Similarity 71.4%; Pred. No. 45;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSKL 7  
 ||:||||  
 Db 43 TRIFSKL 49

# RESULT 7

T32314  
 hypothetical protein F31F4.14 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jul-2000  
 C:Accession: T32314  
 R:Blanchard, M.; Kramer, J.; Elliott, G.; Twyman, B.  
 submitted to the EMBL Data Library, September 1997  
 A:Description: The sequence of C. elegans cosmid F31F4.  
 A:Reference number: Z21149  
 A:Accession: T32314  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-352 <BLA>

A:Cross-references: EMBL:AF024503; PIDN:AA870390.1; GSPDB:G000023; CESP:F31F4.14  
A:Experimental source: strain Bristol NZ; Clone F31F4  
C:Genetics:  
A:Gene: CESP:F31F4.14  
A:Map position: 5  
A:Mutons: 100/3; 170/3; 223/3; 293/3  
C:Superfamily: Caenorhabditis elegans hypothetical protein C33A12.9b

Query Match:	87.9%;	Score 29;	DB 2;	Length 352;
Best Local Similarity:	85.7%;	Pred. No. 55;		
Matches	6;	Conservative	0;	Indels 0;
		Mismatches	0;	Gaps 0;
OY	1	TRIPSKL 7		
	:			
db	48	SRIPSKL 54		

```

RESULT      8
E90089
hypothetical protein orf590 [imported] - Gulliardia theta nucleomorph
C:Species: nucleomorph Gulliardia theta
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: E90089
R:Donaghy, S.; Zanner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Re
Nature 410, 1091-1096, 2001
A:Title: The highly reduced genome of an enslaved algal nucleus.
A:Reference number: A99082; M0ID:11323671
A:Accession: E90089
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-590 <DOU>
A:Cross-references: GB:AF165818; NID:g13794473; PIDN:NAK39848.1; GSPDB:GN00150
C:Genetics:
A:Gene: orf590
A:Map position: 1
A:Genome: nucleomorph
C:Keywords: nucleomorph

```

Query Match	87.9%;	Score 29;	DB 2;	Length 590;
Best Local Similarity	85.7%;	Pred. No. 92;		
Matches	6;	Conservative 1;	Mismatches 0;	Gaps 0;
QY	1	TRIPSKL 7		
	:			
Db	203	SRIPSKL 209		

RESULT 9  
S42220  
transcription initiation factor IID chain p62 - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C:Date: 20-Oct-1994 #sequence\_revision 12-Apr-1996 #text\_change 17-Nov-2000  
C:Accession: S42220  
R:Kokubo, T.; Gouri, D. W.; Wootton, J. C.; Horikoshi, M.; Roeder, R. G.; Nakatani, Y.  
Nucleic 367, 484-487, 1994  
A:Title: Molecular cloning of *Drosophila* TFIID subunits.  
A:Reference number: S42220; MUID:94150630  
A:Accession: S42220  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-592 <KOK>  
A:Cross-references: EMBL:U06459; NID:9458681; PIDN:AN046480.1; PTD:9458682  
C:Genetics:  
A:Gene: FLYBase:1af40  
A:Cross-references: FLYBase:FBgn0000617  
C:Keywords: transcription initiation

Query Match	87.9%;	Score 29;	DB 2;	Length 592;
Best Local Similarity	100.0%;	Pred. No. 92;		

	Matches	0;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	TRFSK	6							
Db	337	TRFSK	342							

RESULT 10  
T18597  
hypothetical protein AC3.5 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18597  
R:McMurray, A  
submitted to the EMBL Data Library, April 1996  
A:Accession: T18597  
A:Reference number: Z18995  
A:Status: preliminary; translated from GB/EMBL/DDBU  
A:Molecule type: DNA  
A:Residues: 1-1071 <M1>  
A:Cross-references: EMBL:Z71177; PIDD:CAM94872.1; GSPDB:GN00023; CESP:AC3.5  
A:Experimental Source: Clone AC3  
C:Genetics:  
A:gene: CESP:AC3.5  
A:Map position: 5  
A:Insertions: 19/3; 124/2; 164/2; 222/3; 376/2; 408/3; 449/1; 537/3; 604/3; 700/2; 808/3

```

Query Match      87.9%; Score 29; DB 2; Length 1071;
Best Local Similarity 85.7%; Pred No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0.

OY      1 TRIPSKL 7
         |||||
db       788 TRIRKRL 794

```

RESULT 11  
 H81081  
 hypothetical protein NMB1455 [imported] - Neisseria meningitidis (strain MC58 serogroup C)  
 C:Species: Neisseria meningitidis  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
 R:Accession: H81081  
 R:Retelling: H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Kiley, G.R.; Hatt, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. et al.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Pizza, M. Science 287, 1809-1815, 2000  
 A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.  
 A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
 A:Reference number: AB1000; MUID:20175755  
 A:Accession: H81081  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-39 <TEXT>  
 A:Cross-references: GB:AE002495; GB:AE002098; NID:g7226690; PIDDN:AA41814.1; PID:g722  
 A:Experimental source: serogroup B, strain MC58  
 C:Genetics:  
 A:Gene: NMB1455

Query Match	84.8%;	Score 28;	DB 2;	length 39;
Best Local Similarity	100.0%;	Pred. No. 10;		
Matches	6;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
			0;	

RESULT 12  
B75636  
transcription regulator, *gerE* family - *Deinococcus radiodurans* (strain R1)  
C:Species: *Deinococcus radiodurans*

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
 C:Accession: B75636  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 S.; Shen, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Ullrich, T.; Zalewski, C.; Ma  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.  
 A:Reference number: A75250; MUID:20036896  
 A:Accession: B75636  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 11179 <NR1>  
 A:Cross-references: GB:AE001827; MID:96460959; PIDN:AAFI2666.1; PID:96460963; TIGR:DRC00  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DRC0012  
 A:Map position: plasmid  
 A:Genome: plasmid  
 A:Note: plasmid CPl

Query Match 84.8%; Score 28; DB 2; Length 179;  
 Best Local Similarity 71.4%; Pred. No. 48;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TRIFSKL 7  
 ||:||||  
 Db 150 TRVFSK 156

## RESULT 13

hypothetical protein BB0776 - Lyme disease spirochete  
 C:Species: *Borrelia burgdorferi* (Lyme disease spirochete)  
 C>Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
 C:Accession: G70196  
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
 Nature 390, 580-586, 1997  
 A:Authors: Smith, H.O.; Venter, J.C.  
 A:Title: Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.  
 A:Reference number: A70100; MUID:98065943  
 A:Accession: G70196  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-187 <NR1>  
 A:Cross-references: GB:AE001177; GB:AE000783; MID:92668711; PIDN:AAC67137.1; PID:9266872  
 A:Experimental source: strain B31

Query Match 84.8%; Score 28; DB 2; Length 187;  
 Best Local Similarity 83.3%; Pred. No. 50;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSK 6  
 ||:||||  
 Db 3 TRVFSK 8

## RESULT 14

hypothetical protein At2g43730 [imported] - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: G84869  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
 euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
 A:Reference number: AB4420; MUID:20083487  
 A:Accession: G84869

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-193 <STO>  
 A:Cross-references: GB:AE002093; MID:92281097; PIDN:AAB64033.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g43730  
 A:Map position: 2

Query Match 84.8%; Score 28; DB 2; Length 193;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RIFSKL 7  
 ||||||  
 Db 173 RIFSKL 178

## RESULT 15

protein R1G11.11 [imported] - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
 C:Accession: E88930  
 R:anonymous, The C. elegans Sequencing Consortium.  
 Science 282, 2012-2018, 1998  
 A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating bio  
 A:Reference number: A75000; MUID:99069613; PMID:9851916  
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C-  
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999.  
 A:Accession: E88930  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-328 <STO>  
 A:Cross-references: GB:chr\_V; PIDN:AAC69085.1; PID:92384860; GSPDB:GN00023; CESP:R1G1  
 C:Genetics:  
 A:Gene: R1G11.11  
 A:Map position: 5

Query Match 84.8%; Score 28; DB 2; Length 328;  
 Best Local Similarity 85.7%; Pred. No. 87;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TRIFSKL 7  
 ||||||  
 Db 280 TRIFSKL 286

Search completed: July 15, 2002, 13:00:59  
 Job time: 465 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: July 15, 2002, 13:23:30 ; Search time 44.9 Seconds

(without alignments)  
6.036 Million cell updates/sec

Title: US-09-712-819a-2

Perfect score: 33

Sequence: 1 TRISKL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	93.9	141	1 HBB_DASAK	P56692 dasyatis ak
2	30	90.9	1237	1 RPA2_SCHPO	099738 schizosacch
3	29	87.9	592	1 T2D5_DROME	P49847 drosophila
4	28	84.8	343	1 ILVC_BUCKL	P49436 drosophila
5	28	84.8	381	1 NY2R_HUMAN	P49146 homo sapien
6	28	84.8	381	1 NY2R_MACMU	O99K74 macaca mula
7	28	84.8	384	1 NY2R_BOVIN	P79113 bos taurus
8	28	84.8	385	1 NY2R_MOUSE	P77235 mus musculu
9	28	84.8	438	1 IAP1_DROME	O24306 drosophila
10	28	84.8	483	1 XYUB_KLEPN	P29444 klebsiella
11	28	84.8	976	1 SCPI_HUMAN	O15431 homo sapien
12	28	84.8	1317	1 N14F_YEAST	P49687 saccharomyc
13	27	81.8	184	1 CLIP_MOUSE	O99738 mus musculu
14	27	81.8	204	1 YC28_RICPR	P41078 rickettsia
15	27	81.8	237	1 YC28_PORPU	P51342 porphyra pu
16	27	81.8	338	1 LDM_HUMAN	P51884 homo sapien
17	27	81.8	343	1 LDM_CHICK	P51890 gallus galli
18	27	81.8	463	1 AROB_CORGL	O46065 corynebacte
19	27	81.8	584	1 COAB_HUMAN	P07295 homo sapien
20	26	78.8	203	1 YEVE_YEAST	P40080 saccharomyc
21	26	78.8	208	1 YC29_GUTTH	O78417 galliardia
22	26	78.8	217	1 GTM1_RAT	P04905 rattus norv
23	26	78.8	369	1 HIS8_ZYMO	P34037 zymomonas m
24	26	78.8	389	1 CD34_CANFA	O28270 canis famli
25	26	78.8	396	1 PPS4_DROME	O28270 canis famli
26	26	78.8	413	1 HP55_TAMSI	O09055 tamias sibi
27	26	78.8	449	1 SR54_RICPR	O99K74 macaca mula
28	26	78.8	459	1 TRPB_PYRHO	O59265 pyrococcus
29	26	78.8	476	1 CHK1_HUMAN	O14757 homo sapien
30	26	78.8	476	1 CHK1_MOUSE	O32280 mus musculu
31	26	78.8	479	1 ACH9_HUMAN	O99K74 macaca mula
32	26	78.8	479	1 ACH9_RAT	P41444 rattus norv
33	26	78.8	580	1 P69_MYCHR	P15362 mycoplasma

34	26	78.8	607	1 UVRG_PSEFL	P32966 pseudomonas
35	26	78.8	622	1 VERN_DROME	O94918 drosophila
36	26	78.8	638	1 YDEC_SCHPO	O10445 schizosacch
37	26	78.8	711	1 FRE2_YEAST	P36035 saccharomyc
38	26	78.8	891	1 DP01_HELPY	P56105 helicobacte
39	26	78.8	897	1 DP01_HELPY	O92359 helicobacte
40	26	78.8	946	1 Y1A2_YEAST	P40559 saccharomyc
41	26	78.8	946	1 Y1A2_YEAST	P29982 murine coro
42	26	78.8	946	1 Y1A2_YEAST	O62059 mus musculu
43	26	78.8	946	1 Y1A2_YEAST	O39575 chlamydomon
44	26	78.8	946	1 Y1A2_YEAST	P21045 vaccinia vi
45	26	78.8	946	1 Y1A2_YEAST	O52620 proteus vul

## ALIGNMENTS

RESULT 1  
HBB\_DASAK STANDARD: PRT, 141 AA.  
AC P56692;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemoglobin beta chain.  
CN HBB.  
CS Dasyatis akajei (red stingray) (Akajei).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
OC Elasmobranchii; Squalia; Hypnosqualia; Pristioraja; Batoidae;  
OC Myliobatiformes; Myliobatoidei; Dasyatidae; Dasyatis.  
CX NCBI\_Taxid=31902;  
RN [1]  
RP SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
RC TISSUE=Blood;  
RA Chong K.T., Miyazaki G., Morimoto H., Oda Y., Park S., Tsukihara T.;  
RT "The three dimensional structures of deoxy and CO forms of hemoglobin  
from Dasyatis akajei, a cartilaginous fish.";  
RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: INVOLVED IN OXYGEN TRANSPORT FROM THE LUNG TO THE  
VARIOUS PERIPHERAL TISSUES.  
CC -1- SUBUNIT: HEMOTETRAMER OF TWO ALPHA CHAINS AND TWO BETA CHAINS.  
CC -1- TISSUE SPECIFICITY: RED BLOOD CELLS.  
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: AB023723; BA075250.1; -.  
CC PDB: 1CG5; 01-APR-99.  
CC PDB: 1CG8; 01-APR-99.  
CC InterPro: IPR000237; Beta\_haem.  
CC InterPro: IPR000971; Globin.  
CC Pfam: PF00042; globin.1.  
CC PRINTS: PRO0814; BETAHAEM.  
CC PROSITE: PS01033; GLOBIN.1.  
CC Heme; Oxygen transport; Transport; Erythrocyte;  
KW 3D-structure.  
FT INIT\_MET 0  
FT METAL 59 59 IRON (HEME DISTAL LIGAND).  
FT METAL 88 88 IRON (HEME PROXIMAL LIGAND).  
SQ SEQUENCE 141 AA; 16289 MW; DAED4F5780AD27B CRC64;  
Query Match 93.9%; Score 31; DB 1; Length 141;  
Best local similarity 85.7%; Pred No. 4.2;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TRISKL 7

```

Db      39 TRFBSKL 45
|||||
RESULT 2
RPA2_SCHPO STANDARD; PRT: 1227 AA.
ID      RPA2_SCHPO
AC      09p7x8;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Probable DNA-directed RNA polymerase I polypeptide 2 (EC 2.7.7.6) (RNA
GN      polymerase I subunit 2).
OS      Schizosaccharomyces pombe (Pisgion yeast).
OC      Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC      Schizosaccharomycetales; Schizosaccharomycetaceae;
OC      Schizosaccharomycetes.
OX      NCBI_TaxID=4896;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=972;
RA      Beck A., Borzym K., Reinhardt R., McDougall R.C., Rajandream M.A.,
RA      Barrell B.G.;
RL      Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC      OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC      SUBSTRATES. RNA POLYMERASE I IS ESSENTIALLY USED TO TRANSCRIBE
CC      RIBOSOMAL DNA UNITS.
CC      -1- CATALYTIC ACTIVITY: N nucleoside triphosphate + N diphosphate +
CC      (RNA)(N).
CC      -1- SUBUNIT: RNA POLYMERASE I CONSISTS OF 14 DIFFERENT SUBUNITS. THIS
CC      SUBUNIT IS THE SECOND LARGEST COMPONENT OF RNA POLYMERASE I (BY
CC      SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC      -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC      FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC      PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC      III FOR 5S AND TRNA GENES.
CC      -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL: AL136535; CAB66435.1;
DR      InterPro: IPR001572; RNA_POL_B.
DR      Pfam: PF00562; RNA_POL_B.1.
DR      PROSITE: PS01166; RNA_POL_BETA.1.
KW      Transferase; DNA-directed RNA polymerase; Transcription; zinc;
KW      zinc-finger; Metal-binding; Nuclear protein.
FT      ZN_FING 1142 1171 C4-type (POTENTIAL).
FT      SQUENCE 1227 AA; 137739 MW; 614A45ADE74D3419 CRC64;
Query Match 90.9%; Score 30; DB 1; Length 1227;
Best Local Similarity 71.4%; Pred. No. 54;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Transcription initiation factor TFIIID 62 kDa subunit (P62) (TAFII-60).
GN      TAF60 OR CG9348.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidae; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX      MEDLINE=94150630; PubMed=7545910;
RA      Kokubo T., Gong D.-W., Wootton J.C., Horikoshi M., Roeder R.G.,
RA      Nakatani Y.;
RT      "Molecular cloning of Drosophila TFIIID subunits.";
RL      Nature 367:484-487(1994).
RN      [2]
RP      SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC      STRAIN=OREGON-R;
RX      MEDLINE=94085406; PubMed=8262073;
RA      Weinzierl R.O., Ruppert S., Dynlacht B.D., Tanese N., Tjian R.;
RT      "Cloning and expression of Drosophila TAFII60 and human TAFII70
RT      reveal conserved interactions with other subunits of TFIIID.";
RL      EMBO J. 12:5303-5309(1993).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BERKELEY;
RX      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers J.-H.C., Blazer G., Chame D., Pfeiffer B.D.,
RA      Wen K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abril J.F., Adyaani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA      Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA      Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA      Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson R., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA      Foster C., Gabrielian A.E., Garg N.S., Gelbair M.M., Glasser K.,
RA      Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA      Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA      Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merklow G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusshorn D.R., Paclet J.M.,
RA      Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA      Splyskas R., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weltschbach J.,
RA      Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA      Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
RA      Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA      Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT      "The genome sequence of Drosophila melanogaster.";
RL      Science 287:2185-2195(2000).
CC      -1- FUNCTION: TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A
CC      CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS
CC      AND REPRESSORS. BINDS TIGHTLY TO TAFII-250 AND ALSO DIRECTLY
CC      INTERACTS WITH TAFII-40.
CC      -1- SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC      NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
CC      -1- SUBCELLULAR LOCATION: Nuclear.

```



```

CC      -1- SIMILARITY: BELONGS TO THE TAF2E FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U06459; AAC6480.1; -.
CC      DR EMBL: L25443; AAA16536.1; -.
CC      DR EMBL: AE003516; AAF49139.1; -.
CC      DR TRANSFAC: T02124; -.
CC      DR FlyBase: FBgn0010417; Tafi60.
CC      DR InterPro: IPR000166; Histone_core.
CC      DR Pfam: PF02969; TAF: 1.
CC      KM Transcription regulation: Nuclear protein.
CC      FT CONFLICT 87 87 F -> L (IN REF. 2).
CC      FT CONFLICT 568 568 E -> D (IN REF. 3).
CC      FT SEQUENCE 592 AA; 64322 MW; 5DBF07AE35E0F7E3 CRC64;
CC
CC      Query Match      87.9%; Score 29; DB 1; Length 592;
CC      Best Local Similarity 100.0%; Pred. No. 46;
CC      Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
Oy      1 TRIPSK 6
Db      337 TRIPSK 342
CC
RESULT 4
ILVC_BUCML STANDARD; PRT; 343 AA.
ID ILVC_BUCML
AC Q9A096;
DC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ketol-acid reductoisomerase (EC 1.1.1.86) (Acetylhydroxy-acid
DE isomeroreductase) (Alpha-keto-beta-hydroxylactil reductoisomerase)
DE (Fragment).
GN ILVC.
OS Buchnera aphidicola (subsp. Macrosiphoniella ludovicianaee).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OC NCBI_TaxID=118105;
RN [1]
RP SEQUENCE FROM N.A.
RA Mernegreen J.J., Moran N.A.;
RT "Accelerated evolutionary rates at biosynthetic loci of Buchnera-
RT urelecon."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: (R)-2,3-dihydroxy-3-methylbutanoate + NADP(+)
CC      - (S)-2-hydroxy-2-methyl-3-oxobutanoate + NADPH.
CC -1- PATHWAY: VALINE AND ISOLEUCINE BIOSYNTHESIS; SECOND STEP.
CC -1- SIMILARITY: BELONGS TO THE KETOL-ACID REDUCTOISOMERASE FAMILY.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: AF217557; AAK01028.1; -.
CC      DR InterPro: IPR000506; Acetylhydrazac_isomrctase.
CC      DR Pfam: PF01450; ILVC; 1.
CC      KM Oxidoreductase; Branched-chain amino acid biosynthesis; NADP.
CC      FT ACT SITE 65 65 POTENTIAL.
CC      FT NON_TER 343 343
CC      SEQUENCE 343 AA; 39012 MW; 67369DC0E240113C CRC64;

```

```

CC      Query Match      84.8%; Score 28; DB 1; Length 343;
CC      Best Local Similarity 100.0%; Pred. No. 46;
CC      Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
Oy      2 RIFSKL 7
Db      227 RIFSKL 232
CC
RESULT 5
ID NY2R_HUMAN STANDARD; PRT; 381 AA.
AC P49146; O13281; O13457; Q9UE67;
DC 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neuropeptide Y receptor type 2 (NPY2-R) (NPY-Y2 receptor).
DE NPY2R.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hippocampus;
RX MEDLINE=96070760; PubMed=7592910;
RX Gerald C., Walker M.W., Vayssé P.J.-J., He C., Branchek T.A.,
RX Weisshank R.L.;
RT "Expression cloning and pharmacological characterization of a human
RT hippocampal neuropeptide Y/peptide YY Y2 receptor subtype."
RL J. Biol. Chem. 270:26758-26761(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96226058; PubMed=8632753;
RX Gehlert D.R., Beavers L.S., Johnson D., Gackenhelmer S.L.,
RX Schober D.A., Gadeki R.A.;
RT "Expression cloning of a human brain neuropeptide Y Y2 receptor."
RL Mol. Pharmacol. 49:224-228(1996).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96032678; PubMed=7559383;
RX Rose P.M., Ferrandes P., Lynch J.S., Frazier S.T., Fisher S.M.,
RX Kodukula K., Kienzie B., Seethala R.;
RT "Cloning and functional expression of a cDNA encoding a human type 2
RT neuropeptide Y receptor."
RL J. Biol. Chem. 270:22661-22664(1995).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=96209788; PubMed=8643460;
RX Yan H., Yang J., Marasco J., Yamaguchi K., Brenner S., Collins F.,
RX Karbon W.;
RT "Cloning and functional expression of cDNAs encoding human and rat
RT proc. Natl. Acad. Sci. U.S.A. 93:4661-4665(1996).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=97131518; PubMed=8975716;
RX Ammer D.A., Eadie D.M., Wong D.J., Ma Y.-Y., Kolakowski L.F. Jr.,
RX Yang-Feng T.L., Thompson D.A.;
RT "Characterization of the human type 2 neuropeptide Y receptor gene
RT (NPY2R) and localization to the chromosome 4q region containing the
RT type 1 neuropeptide Y receptor gene."
RL Genomics 38:392-398(1996).
RN [6]
RP SEQUENCE FROM N.A.
RX Zaslavny R.L.;
RT "Human neuropeptide Y Y2 receptor gene."
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RECEPTOR FOR NEUROPEPTIDE Y AND PEPTIDE YY. THE RANK
CC ORDER OF AFFINITY FOR THIS RECEPTOR FOR PANCREATIC POLYPEPTIDES IS

```

CC	LEU_NPY > PYY (3-36) > NPY (2-36) > [Ile-34] PP >
CC	[Phe-31,Pro-34] PYY > PY [Pro-34] PYY AND NOT FREE ACID.
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein
CC	-1- TISSUE SPECIFICITY: HIGH LEVELS IN AMYGDALA, CORPUS CALLOSUM,
CC	HIPPOCAMPUS AND SUBSTANTIAL NUCLEUS. ALSO DETECTABLE IN CAUDATE
CC	NUCLEUS, HIPHOCALAMUS AND SUBSTANTIA NIGRA.
CC	-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC	HIGHEST TO TACHYKININS RECEPTORS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC	between the Swiss Institute of Bioinformatics and the EMBL collaboration -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sdb.ch">license@isb-sdb.ch</a> ).
CC	-----
DR	EMBL: U36269; AAC50281.1; -
DR	EMBL: U47766; AAB04120.1; -
DR	EMBL: U35500; AAA93170.1; -
DR	EMBL: U42389; AAB07760.1; -
DR	EMBL: U50146; AAC51115.1; -
DR	EMBL: U76254; AAD00248.1; -
DR	GCRDB: GCR_1945; -
DR	GCRDB: GCR_1946; -
DR	GCRDB: GCR_1958; -
DR	GCRDB: GCR_1974; -
DR	GCRDB: GCR_2066; -
DR	MIM: 162642; -
DR	Interpro: IPR00276; GPCR_Rhodopsn.
DR	Pfam: PF00001; 7tm.1.1.
DR	PRINTS: PR00237; GPCRRHODOPSN.
DR	PROSITE: PS00237; G_PROTEIN_RECPE_FL_1; 1.
DR	PROSITE: PS50262; G_PROTEIN_RECPE_FL2; 1.
KM	G-protein coupled receptor; Transmembrane; Glycoprotein;
KM	Phosphorylation; Lipoprotein; Palmitate.
FT	DOMAIN 1 49
FT	TRANSEM 50 72
FT	DOMAIN 73 82
FT	TRANSEM 83 104
FT	DOMAIN 105 124
FT	TRANSEM 125 146
FT	DOMAIN 147 166
FT	TRANSEM 167 187
FT	DOMAIN 188 214
FT	TRANSEM 215 240
FT	DOMAIN 241 268
FT	TRANSEM 269 291
FT	DOMAIN 292 304
FT	TRANSEM 305 328
FT	DOMAIN 329 381
FT	CARBOHYD 11 11
FT	DLSUFID 123 203
FT	LIPID 342 342
FT	CONFLICT 134 134
FT	CONFLICT 172 172
FT	CONFLICT 174 174
FT	CONFLICT 202 202
FT	CONFLICT 209 209
FT	CONFLICT 248 248
FT	CONFLICT 311 311
QO	SEQUENCE 381 AA: 42731 MW: 70016CD169597BC7 CRC64:

Query Match	84.88;	Score 28;	DB 1;	Length 381;
Best Local Similarity	85.78;	Pred. NO. 51;		
Matches	6;	Conservative	1;	Mismatches 0;
				Indels 0;
				Gaps 0;

OY	1	TRISKL	7
Db	240	TRISKL	246

RESULT	5				
NY2R_MACMU		STANDARD:	PRT:	361 AA.	
AC	O96K74;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Neuropeptide Y receptor type 2 (NPY-2) (NPY-Y2 receptor).				
GN	NPY2R.				
OS	Macaca mulatta (Rhesus macaque).				
OC	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;				
OC	Cercopitheciinae; Macaca.				
OX	NCBI_TaxID=9544;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-1184974; PubMed-11287088;				
RA	Gheltier D.R., Yang P., George C., Wang Y., Schober D., Baez M.;				
RA	Gaekhelmeier S., Johnson D., Beavers L.S., Gauski R.A., Baez M.;				
RT	"Cloning and characterization of Rhesus monkey neuropeptide Y receptor				
RT	subtypes.";				
RL	Peptides 22:345-350(2001).				
CC	-1- FUNCTION: RECEPTOR FOR NEUROPEPTIDE Y AND PEPTIDE YY.				
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.				
CC	-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.				
CC	HIGHEST TO TACHYKININS RECEPTORS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL: AF030390; AAC04072.1; -				
DR	InterPro: IPR000276; GPCR_Rhodpsn.				
DR	Pfam: PF00001; 7tm_1; 1.				
DR	PRINTS: PR00237; GPCR_RHODPSN.				
DR	PROSITE: PS00237; G_PROTEIN_REC_P1_1; 1.				
DR	PROSITE: PS00262; G_PROTEIN_REC_P1_2; 1.				
KW	G-protein coupled receptor; Transmembrane; Glycoprotein;				
KW	Phosphorylation; Lipoprotein; Palmitate.				
FT	DOMAIN	1	49		EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	50	72		1 (POTENTIAL).
FT	DOMAIN	73	82		CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	83	104		2 (POTENTIAL).
FT	DOMAIN	105	124		EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	125	146		3 (POTENTIAL).
FT	DOMAIN	147	166		CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	167	187		4 (POTENTIAL).
FT	DOMAIN	188	214		EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	215	240		5 (POTENTIAL).
FT	DOMAIN	241	268		CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	269	291		6 (POTENTIAL).
FT	DOMAIN	292	304		EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	305	328		7 (POTENTIAL).
FT	CARBOHYD	329	381		CYTOPLASMIC (POTENTIAL).
FT	DISULFID	123	203		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	LIPID	342	342		BY SIMILARITY.
QO	SEQUENCE	361 AA;	42729 MW;		PALMITATE (POTENTIAL).
					DB89691DCC05A49D CRC64;

Query Match	84.8%;	Score 28;	DB 1;	Length 381;
Best Local Similarity	85.7%;	Pred. No. 51;		
Matches	6;	Conservative	1;	Mismatches
			0;	Indels
			0;	Gaps
			0;	

OY	1	TRIESKL	7
		111:111	
Db	240	TRIWSKL	246

```

RESULT 7
NY2R_BOVIN STANDARD: PRT: 384 AA.
ID NY2R_BOVIN
P79113;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Neuropeptide Y receptor type 2 (NPY2-R) (NPY-Y2 receptor).
GN NPY2R.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Kuminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Amar D.A., Kolakowski L.F. Jr., Eagle D.M., Wong D.J., Ma Y.Y.,
RA Yang-Feng T.L., Thompson D.A.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: RECEPTOR FOR NEUROPEPTIDE Y AND PEPTIDE YY.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC HIGHEST TO TACHYKININS RECEPTORS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: U50144; AAB40600.1;
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam: PF00001; 7tm.1; 1.
DR PRINTS: PR00237; GPCR_RHODOPSN.
DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Phosphorylation; Lipoprotein; Palmitate.
FT DOMAIN 1 52 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 53 75 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 76 85 2 (POTENTIAL).
FT TRANSSEM 86 107 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 108 127 3 (POTENTIAL).
FT TRANSSEM 128 149 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 150 169 4 (POTENTIAL).
FT TRANSSEM 170 190 5 (POTENTIAL).
FT DOMAIN 191 217 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 218 243 5 (POTENTIAL).
FT DOMAIN 244 271 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 272 294 6 (POTENTIAL).
FT DOMAIN 295 307 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 308 331 7 (POTENTIAL).
FT DOMAIN 332 384 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 13 13 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DISULFID 126 206 BY SIMILARITY.
FT LIPID 345 345 PALMITATE (POTENTIAL).
SQ SEQUENCE 384 AA; 42943 MW; 468D19CBA8F296B1 CRC64;

```

Query Match 84.8%; Score 28; DB 1; Length 384;  
 Best Local Similarity 85.7%; Pred. No. 51;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TRIPSKL 7  
 |||:||||  
 Db 243 TRIPSKL 249

RESULT 8  
 NY2R\_MOUSE STANDARD: PRT: 385 AA.

```

AC p97295;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Neuropeptide Y receptor type 2 (NPY2-R) (NPY-Y2 receptor).
GN NPY2R.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Brain;
RX MEDLINE=97071731; PubMed=8914576;
RA Nakamura M., Aoki Y., Hirano D.;
RT "Cloning and functional expression of a cDNA encoding a mouse type 2
RT neuropeptide Y receptor."
RL Biochim. Biophys. Acta 1284:134-137(1996).
CC -!- FUNCTION: RECEPTOR FOR NEUROPEPTIDE Y AND PEPTIDE YY.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC HIGHEST TO TACHYKININS RECEPTORS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: D86238; BAA13050.1;
DR GCRDB: GCR_1146;
DR MGD: MGI:108418; NPY2r.
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam: PF00001; 7tm.1; 1.
DR PRINTS: PR00237; GPCR_RHODOPSN.
DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Phosphorylation; Lipoprotein; Palmitate.
FT DOMAIN 1 53 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 54 76 1 (POTENTIAL).
FT DOMAIN 77 86 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 87 108 2 (POTENTIAL).
FT DOMAIN 109 128 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 129 150 3 (POTENTIAL).
FT DOMAIN 151 170 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 171 191 4 (POTENTIAL).
FT DOMAIN 192 218 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 219 244 5 (POTENTIAL).
FT DOMAIN 245 272 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 273 295 6 (POTENTIAL).
FT DOMAIN 296 308 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 309 332 7 (POTENTIAL).
FT DOMAIN 333 385 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 15 15 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DISULFID 127 207 BY SIMILARITY.
FT LIPID 346 346 PALMITATE (POTENTIAL).
SQ SEQUENCE 385 AA; 43099 MW; 5AD7E4C9B7077085 CRC64;

```

Query Match 84.8%; Score 28; DB 1; Length 385;  
 Best Local Similarity 85.7%; Pred. No. 51;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TRIPSKL 7  
 |||:||||  
 Db 244 TRIPSKL 250

RESULT 9  
 IAP1\_DROME

```

ID   IAP1.DROME          STANDARD:      PRT:   438 AA.
DC   024306;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, last sequence update)
DT   16-OCT-2001 (Rel. 40, last annotation update)
DE   Apoptosis 1 inhibitor (inhibitor of apoptosis 1) (DIAP1) (Thread
DE   protein).
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Empidoidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE-EYE Imaginal disk;
RX   MEDLINE=96128128; PubMed=8548811;
RA   Hay B.A., Messerman D.A., Rubin G.M.;
RT   "Drosophila homologs of baculovirus inhibitor of apoptosis proteins
RT   function to block cell death.";
RL   Cell 83:1253-1262(1995).
CC   -1- FUNCTION: APOPTOTIC SUPPRESSOR. OVEREXPRESSION SUPPRESSES RPR AND
CC   -1- HD-DEPENDENT CELL DEATH IN THE EYE.
CC   -1- SIMILARITY: BELONGS TO THE IAP FAMILY.
CC   -1- SIMILARITY: CONTAINS 2 BIR REPEATS.
CC   -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL: I49440; AAC41609.1; -
DR   HSSP: Q13490; 10BH.
DR   Flybase: FBgn0003691; th.
DR   InterPro: IPR001370; BIR.
DR   InterPro: IPR001841; Znf_ring.
DR   Pfam: PF00653; BIR_2.
DR   Pfam: PF00097; zf-C3HC4; 1.
DR   SMART: SM00238; BIR_2.
DR   SMART: SM00184; RING_1.
DR   PROSITE: PS01282; BIR_REPEAT_1; 2.
DR   PROSITE: PS0143; BIR_REPEAT_2; 2.
DR   PROSITE: PS00518; zf_RING_1; FALSE_NEG.
DR   PROSITE: PS0089; zf_RING_2; 1.
KW   Apoptosis; zinc-finger; Repeat.
FT   REPEAT            44      110      BIR 1.
FT   REPEAT            225     293      BIR 2.
FT   ZN_FING           391     426      RING-TYPE.
SQ   SEQUENCE   438 AA; 48098 MW; AQC22C8EDF5AEF29 CRC64;

```

```

Query Match      84.8%; Score 28; DB 1; Length 438;
Best Local Similarity 71.4%; Pred. No. 58;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

OY   1 TRIFSKL 7
     |||||:
DB   360 TRIFPKI 366

```

```

RESULT 10
XLIB_KLEPN
ID   XLIB_KLEPN          STANDARD:      PRT:   463 AA.
AC   P29444;
DT   01-APR-1993 (Rel. 25, Created)
DT   03-APR-1993 (Rel. 25, last sequence update)
DT   30-MAY-2000 (Rel. 39, last annotation update)
DE   Xylulose kinase (EC 2.7.1.17) (Xylulokinase).
XLIB_

```

```

OS   Klebsiella pneumoniae.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Klebsiella.
OX   NCBI_TaxID=573;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1033-5P14 / KAY2026;
RX   MEDLINE=92374988; PubMed=1324398;
RA   Feldmann S.D., Salm H., Sprenger G.A.;
RT   "Cloning and expression of the genes for xylulose isomerase and
RT   xylulokinase from Klebsiella pneumoniae 1033 in Escherichia coli
RT   K12.";
RL   Mol. Genet. 234:201-210(1992).
CC   -1- CATALYTIC ACTIVITY: ATP + D-xylulose -> ADP + D-xylulose 5-
CC   phosphate.
CC   -1- SIMILARITY: BELONGS TO THE FRUCTOKINASE / GLUCOKINASE /
CC   GLYCEROKINASE / XYLULOXINASE FAMILY.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL: X61059; CAA43390.1; -
DR   PIR: S25070; S25070.
DR   InterPro: IPR000577; FGXY_kin.
DR   Pfam: PF00370; FGXY_1.
DR   Pfam: PF02782; FGXY_C; 1.
DR   PROSITE: PS00445; FGXY_KINASES_2; 1.
DR   PROSITE: PS00933; FGXY_KINASES_1; 1.
KW   Transferase; Kinase; Xylulose metabolism.
SQ   SEQUENCE   463 AA; 51907 MW; 835FE49401BF872 CRC64;

```

```

Query Match      84.8%; Score 28; DB 1; Length 483;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY   2 RIRSKL 7
     |||||
DB   473 RIRSKL 478

```

```

RESULT 11
SCPL_HUMAN
ID   SCPL_HUMAN          STANDARD:      PRT:   976 AA.
AC   015431; 014963;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, last sequence update)
DT   16-OCT-2001 (Rel. 40, last annotation update)
DE   Synaptonemal complex protein 1 (SCP-1 protein).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE-Testis;
RX   MEDLINE=97224467; PubMed=9119375;
RA   Neuwissen R.L.J., Meerts I., Hoovers J.M.N., Teschot N.J.,
RA   Heyling C.;
RT   "Human synaptonemal complex protein 1 (SCP1): isolation and
RT   characterization of the cDNA and chromosomal localization of the
RT   gene.";
RL   Genomics 39:377-384(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE-Testis;
RX   MEDLINE=98037449; PubMed=9371398;

```

RA Kondoh N., Nishina Y., Tsuchida J., Koga M., Tanaka H., Uchida K.,  
 RA Inazawa J., Takeo M., Nozaki M., Nojima H., Matsumaya K., Nemiki M.,  
 RA Okuyama A., Nishimune Y.;  
 RT "Assignment of synaptonemal complex protein 1 (SCP1) to human  
 RT chromosome 1p13 by fluorescence in situ hybridization and its  
 RT expression in the testis.";  
 RL Cytogenet. Cell Genet. 78:103-104(1997).  
 CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF  
 CC SYNAPTONEMAL COMPLEXES (SCS). FORMED BETWEEN HOMOLOGOUS  
 CC CHROMOSOMES DURING MEIOTIC PROPHASE.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF  
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.  
 CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS  
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE  
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE  
 CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: TESTIS.  
 CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,  
 CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL  
 CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; X95654; CAA64956.1; -;  
 DR EMBL; D67035; BAA22586.1; -;  
 DR MIM; 602162; -;  
 KW Nuclear protein; Meiosis; Cell division; Phosphorylation;  
 KM DNA-binding; Coiled coil.  
 FT DOMAIN 12 100 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 107 798 COILED COIL (POTENTIAL).  
 FT DOMAIN 117 120 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 679 682 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 880 883 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 961 969 ARG/LYS-RICH (BASIC).  
 FT DOMAIN 46 46 L->F (IN REF. 2).  
 FT CONFLICT 106 106 F->Y (IN REF. 2).  
 FT CONFLICT 153 153 F->C (IN REF. 2).  
 FT CONFLICT 161 161 K->T (IN REF. 2).  
 FT CONFLICT 168 168 E->D (IN REF. 2).  
 FT CONFLICT 216 216 N->S (IN REF. 2).  
 FT CONFLICT 225 226 HG->FE (IN REF. 2).  
 FT CONFLICT 350 350 K->N (IN REF. 2).  
 FT CONFLICT 360 360 E->D (IN REF. 2).  
 FT CONFLICT 400 401 KN->NY (IN REF. 2).  
 FT CONFLICT 406 406 K->I (IN REF. 2).  
 FT CONFLICT 415 415 K->T (IN REF. 2).  
 FT CONFLICT 449 449 E->D (IN REF. 2).  
 FT CONFLICT 483 510 IOLTAITTSSEQYSEYKREKDKTELENEK -> YSYCHYHKW  
 FT CONFLICT 510 510 TVLPKRGPRKLSKREK (IN REF. 2).  
 FT CONFLICT 516 528 LSHCNKLSLENEK -> YFTLOQASPPPN (IN REF. 2).  
 FT CONFLICT 549 549 N->I (IN REF. 2).  
 FT CONFLICT 560 560 K->T (IN REF. 2).  
 FT CONFLICT 805 805 E->D (IN REF. 2).  
 FT CONFLICT 941 941 P->S (IN REF. 2).  
 SO SEQUENCE 976 AA; 114069 MW; 8BA81D042AC2696B CRC64;

Query Match 84.8%; Score 28; DB 1; Length 976;  
 Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIRSKL 7  
 DB 103 SRVESKL 109

RESULT 12  
 ID N145\_YEAST STANDARD; PRT; 1317 AA.  
 AC P49687;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Nucleoporin NUP145 (Nuclear pore protein NUP145).  
 GN NUP145 OR YGL092W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Saccharomycetes: Ascomycota: Saccharomycotina: Saccharomycetes:  
 OC Eukaryota: Fungi; Ascomycota: Saccharomycetales: Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94320139; PubMed=8044840;  
 RA Fabre E., Boelens W.C., Wimmer C., Mattaj J.W., Hurt E.C.;  
 RT "Nup145p is required for nuclear export of mRNA and binds  
 RT homopolymeric RNA in vitro via a novel conserved motif.";  
 RL Cell 78:275-289(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 26109 / X2180;  
 RX MEDLINE=94253246; PubMed=8195299;  
 RA Wente S.R., Blobel G.;  
 RT "NUP145 encodes a novel yeast glycine-leucine-phenylalanine-glycine  
 RT (GLRF) nucleoporin required for nuclear envelope structure.";  
 RL J. Cell Biol. 125:955-969(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=S288C;  
 RX MEDLINE=97435481; PubMed=929012;  
 RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;  
 RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae  
 RT chromosome VII.";  
 RL Yeast 13:1077-1090(1997).  
 CC -1- FUNCTION: MAY PLAY A DIRECT ROLE IN NUCLEOCYTOPLASMIC EXPORT OF  
 CC RNA AND ALSO IN PROTEIN IMPORT. BINDS HOMOPOLYMERIC RNA.  
 CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.  
 CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.  
 CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; X76557; CAA54057.1; -;  
 DR EMBL; Z32672; CAA83584.1; -;  
 DR EMBL; Z72614; CAA86798.1; -;  
 DR SGD; S0003060; NUP145.  
 KW Nuclear protein; Transport; Repeat; RNA-binding.  
 FT CONFLICT 281 282 NA->OR (IN REF. 2).  
 FT CONFLICT 1142 1142 L->S (IN REF. 2).  
 FT CONFLICT 1310 1317 LMKCTYKI -> FEYVI (IN REF. 2).  
 SO SEQUENCE 1317 AA; 145660 MW; 59399D86B53030 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 1317;  
 Best Local Similarity 83.3%; Pred. No. 1.0e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIRSK 6  
 DB 1068 TRVPSK 1073  
 RESULT 13  
 ID C1IP\_MOUSE STANDARD; PRT; 184 AA.

```

AC O9CR36: Q9D7K7: O9CTZ5;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE CAIL protein homolog.
GN CAIL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Stomach, and Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kodera K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guslinich S., Hill D., Hoffmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Monbaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski B., Yoshida K., Hasegawa Y., Kawaji H., Kohlsaki S.,
RA Hayashizaki Y.;
RL "Functional annotation of a full-length mouse cDNA collection.";
NT Nature 409:685-690(2001).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AK008990; BAB26010.1; -
DR EMBL: AK008622; BAB25784.1; -
DR EMBL: AK008641; BAB25801.1; -
DR EMBL: AK008647; BAB25805.1; -
DR EMBL: AK008722; BAB25856.1; -
DR EMBL: AK008745; BAB25872.1; -
DR EMBL: AK008933; BAB25973.1; -
DR EMBL: AK008956; BAB25988.1; -
DR EMBL: AK009145; BAB26103.1; -
DR EMBL: AK019050; BAB31525.1; -
FT CONFLICT 113 113 P -> L (IN REF. 1: BAB26103).
SQ SEQUENCE 184 AA; 20134 MW; 288982F0404EFA8B CRC64;

```

```

Query Match      81.8%; Score 27; DB 1; Length 184;
Best Local Similarity 83.3%; Pred. No. 43;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 TRIFSK 6
DB 73 TRIFSK 78

```

```

RESULT 14
Y860_RICPR STANDARD; PRT; 204 AA.
AC P1078;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Hypothetical protein RP860.
GN RP860.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxId=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RA Wood D.O., Marks G.L., Winkler H.H.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Anderson S.G.E., Zomrodipour A., Andersson J.O.,
RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Neeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U02878; AAB81402.1; -
DR EMBL: AJ235273; CA15284.1; -
DR InterPro: IPR002741; DUF56.
DR Pfam: PF01879; DUF56; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 204 AA; 23196 MW; E4CE95B478BCCCA6 CRC64;

```

```

Query Match      81.8%; Score 27; DB 1; Length 204;
Best Local Similarity 71.4%; Pred. No. 47;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 TRIFSK 7
DB 63 TRIFSK 69

```

```

RESULT 15
YC28_PORPU STANDARD; PRT; 237 AA.
AC P51342;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical transcriptional regulator yct28.
GN YCF28.
OS Porphyra purpurea.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCBI_TaxId=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AVONPORT;
RA Keith M.E., Munholland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RL genome.";
RL Plant Mol. Biol. Rep. 13:333-335(1995).
CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).

CC -----  
DR EMBL; U38804; AAC08228.1; -  
DR InterPro; IPR001808; HTH\_CRP.  
DR InterPro; IPR000595; CNMP\_binding.  
DR Pfam; PF00027; CNMP\_binding; 1.  
DR Pfam; PF00325; crp; 1.  
DR SMART; SM00419; HTH\_CRP; 1.  
DR SMART; SM00100; CNMP; 1.  
DR PROSITE; PS00042; HTH\_CRP\_FAMILY; 1.  
KW Hypothetical protein; Transcription regulation; DNA-binding;  
FT Chloroplast.  
FT DNA\_BIND 188 207 H-T-H MOTIF (BY SIMILARITY).  
SQ SEQUENCE `237 AA: 27373 MW: CD8AFA7C1DD4F333 CRC64;

Query Match 81.8%; Score 27; DB 1; Length 237;  
Best Local Similarity 85.7%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TRIISKL 7  
||| |||  
Db 203 TRIISKL 209

Search completed: July 15, 2002, 13:23:31  
Job time: 1447 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using SW model

Run on: July 15, 2002, 13:22:39 ; Search time 172.49 Seconds  
(without alignments)  
7.020 Million cell updates/sec

Title: US-09-712-819A-2

Perfect score: 33

Sequence: 1 TRIPSKL 7

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL.19:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.mhc:\*  
9: sp.organelle:\*  
10: sp.phage:\*  
11: sp.plant:\*  
12: sp.podent:\*  
13: sp.virus:\*  
14: sp.vertibrate:\*  
15: sp.unclassified:\*  
16: sp.virus:\*  
17: sp.bacteriaph:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	97.0	529	5	016232
2	30	90.9	499	5	016230
3	30	90.9	1106	16	09CBP3
4	29	87.9	220	11	09CV09
5	29	87.9	244	16	09ABK0
6	29	87.9	291	16	097KW7
7	29	87.9	297	4	09BSH4
8	29	87.9	352	5	017134
9	29	87.9	419	5	09BMX4
10	29	87.9	457	2	093U20
11	29	87.9	579	10	09FN09
12	29	87.9	579	10	093Y14
13	29	87.9	590	10	096RV2
14	29	87.9	606	5	0961B6
15	29	87.9	1071	5	017405
16	29	87.9	1522	10	0942K3

17	28	84.8	39	16	09JYS2	09JYS2 neisseria m
18	28	84.8	76	11	070424	070424 rattus norv
19	28	84.8	149	2	09XBC4	09XBC4 anycolatops
20	28	84.8	179	16	09R2H3	09R2H3 delnoccocus
21	28	84.8	181	2	09APM6	09APM6 pseudocall
22	28	84.8	187	16	051717	051717 borrelia bu
23	28	84.8	193	10	022830	022830 arabidopsis
24	28	84.8	295	16	098KJ9	098KJ9 rhizobium l
25	28	84.8	328	5	016953	016953 caenorhabdi
26	28	84.8	381	11	09EKC0	09EKC0 rattus norv
27	28	84.8	422	5	020082	020082 caenorhabdi
28	28	84.8	438	5	09VUX5	09VUX5 drosophila
29	28	84.8	545	12	09MAT6	09MAT6 chuzen viru
30	28	84.8	580	10	0940F9	0940F9 arabidopsis
31	28	84.8	586	10	09SZF2	09SZF2 arabidopsis
32	28	84.8	691	13	09PRH5	09PRH5 oncorhynch
33	28	84.8	702	4	09HOK9	09HOK9 homo sapien
34	28	84.8	744	10	048589	048589 arabidopsis
35	28	84.8	834	10	022187	022187 arabidopsis
36	28	84.8	1304	5	09BM81	09BM81 clonorchis
37	28	84.8	1304	5	09BM79	09BM79 clonorchis
38	28	84.8	1311	5	09BM80	09BM80 clonorchis
39	28	84.8	1491	5	09NBD7	09NBD7 drosophila
40	28	84.8	1492	5	09NLS7	09NLS7 drosophila
41	27	81.8	156	17	029614	029614 archaeglob
42	27	81.8	161	11	09P8A9	09P8A9 mus musculu
43	27	81.8	252	12	09DY77	09DY77 putella xy
44	27	81.8	258	4	09H737	09H737 homo sapien
45	27	81.8	338	4	0960M7	0960M7 homo sapien

#### ALIGNMENTS

RESULT 1  
ID 016232 PRELIMINARY: PRT: 529 AA.  
AC 016232;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHEMETICAL 61.2 KDA PROTEIN.  
GN P07G11.3  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae.  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE-99069613; PubMed-9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.",  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Sammons L., Wohlmann P., Sansone J.;  
RT "The sequence of C. elegans cosmid P07G11.",  
RL Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submission.",  
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF016419; AAC24048.1;  
DR InterPro: IPR002875; DUF23.  
DR Pfam: PF01697; DUF23; 1.  
DR Prodom: PD004153; DUF23; 1.  
DR Hypothetical protein.  
SQ SEQUENCE 529 AA; 61221 MW; CEAAFF583E95558B0 CRC64;

Query Match 97.0%; Score 32; DB 5; Length 529;  
Best Local Similarity 85.7%; Pred. No. 27;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSRL 7  
||:||||  
Db 440 TRVSRL 446

RESULT 2  
ID 016230 PRELIMINARY; PRT; 499 AA.

AC 016230; 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 58.2 KDA PROTEIN.  
GN F07G11.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdillida; Rhabdillioidea;  
OC Rhabdillida; Pelodermata; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium. ";  
RL Science 282:2012-2018(1998).  
RM (2)  
RN SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Sammons L., Woldmann P., Sansone J.;  
RT "The sequence of C. elegans cosmid F07G11. ";  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RM (3)  
RN SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submission. ";  
RL EMBL: AF016419; AAG24051.1; -  
DR InterPro: IPR002875; DUF23.  
DR Pfam: PF01697; DUF23. 1.  
DR Prodom: PD004153; DUF23. 1.  
KW Hypothetical protein.  
SQ SEQUENCE 499 AA; 58227 MW; 55F8F7E6871FB9F5 CRC64;

Query Match 90.9%; Score 30; DB 5; Length 499;  
Best Local Similarity 85.7%; Pred. No. 77;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSRL 7  
||:||||  
Db 410 TRVSRL 416

RESULT 3  
ID 09CBP3 PRELIMINARY; PRT; 1106 AA.

AC 09CBP3;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE POSSIBLE TRANSCRIPTIONAL REGULATOR.  
GN ML1753.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1769;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-TN;  
RX MEDLINE=21128732; PubMed=11234002;  
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,  
RA Holtroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrall B.G.;  
RT "Massive gene decay in the leprosy bacillus. ";  
RL Nature 409:1007-1011(2001).  
CC -1 SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
DR EMBL: AL583923; CAC30706.1; -.  
DR Leproma; ML1753; -.  
DR InterPro: IPR000767; Disease\_resist.  
DR InterPro: IPR000792; HTH\_LuxR.  
DR Pfam: PF00196; GntE; 1.  
DR PRINTS: PR00364; DISEASERESIST.  
DR PRINTS: PR00038; HTHLUXR.  
DR SMART: SM00421; HTH\_LuxR; 1.  
KW Complete proteome; DNA-binding; Transcription regulation.  
SQ SEQUENCE 1106 AA; 119374 MW; 75368FB4C7C99DE2 CRC64;

Query Match 90.9%; Score 30; DB 16; Length 1106;  
Best Local Similarity 85.7%; Pred. No. 17+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSRL 7  
||:||||  
Db 1084 TRVSRL 1090

RESULT 4  
ID 09CV09 PRELIMINARY; PRT; 220 AA.

AC 09CV09;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE 2310066118RIK PROTEIN (FRAGMENT).  
GN 2310066118RIK.  
OS Mus musculus (mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J; TISSUE=TONGUE;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi T., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochava H.,  
RA Kuehl P., Lewis S., Matsuo Y., Mikado I., Pesole G., Quackenbush J.,  
RA Schirral L.M., Staudil F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombertis P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez T., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,

RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK010072; BAB26682.1; -.  
 DR MGD: MGI:1917457; 2310066118RLK.  
 DR InterPro: IPR002876; DUF28.  
 DR ProDom: PD004323; DUF28; 1.  
 FT MON\_TER 220 220 MW: 23862 MW: F78AC7088AB4B6 CRC64;  
 SQ SEQUENCE 220 AA; 23862 MW; F78AC7088AB4B6 CRC64;

Query Match 87.9%; Score 29; DB 11; Length 220;  
 Best Local Similarity 85.7%; Pred. No. 59;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSK 7  
 :|||||  
 DB 75 SRIFSKL 81

RESULT 5  
 09ABK0 PRELIMINARY; PRT; 244 AA.  
 AC 09ABK0; (TEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TEMBLrel. 19, Last annotation update)  
 DE HYPOTHEICAL PROTEIN CC1353.  
 GN CC1353.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter  
 OX NCBI\_TaxID=69394;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-AFCC 19089 / CB15;  
 RC MEDLINE-21173698; PubMed-11259647;  
 RX Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 Porocka I., Nelson W.C., Newton A.S., Gwinn M.L., Haft D.H., Ely B.,  
 Dacey R.T., Dodson R.J., Durkin A.S., Matveyev J., Ermolaeva M., White O.,  
 Kolony J.F., Sitt J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 Ueberback T., Tran K., Wolf A., Yamahevan J., Ermolaeva M., White O.,  
 Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of Caulobacter crescentus."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL: AE005810; AAK23334.1; -.  
 DR TIGR: CC1353; -.  
 DR InterPro: IPR001601; Meth-transf.  
 DR InterPro: IPR000051; SAM\_bind.  
 KW Hypothetical protein, Complete proteome.  
 SQ SEQUENCE 244 AA; 26206 MW; DE0382178696DEF1 CRC64;

Query Match 87.9%; Score 29; DB 16; Length 244;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSK 6  
 :|||||  
 DB 69 TRIFSK 74

RESULT 6  
 097KW7 PRELIMINARY; PRT; 291 AA.  
 AC 097KW7;  
 DT 01-OCT-2001 (TEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TEMBLrel. 18, Last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
 DE PHOSPHATIDYLISERINE DECARBOXYLASE.  
 GN CAC0799.  
 OS Clostridium acetobutylicum.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1488;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-AFCC 824 / DSM 792 / VKM B-1787;  
 RC MEDLINE-21359325; PubMed-11466286;  
 RX Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
 Tatusov R.L., Sabatne F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
 RA Hennett G.N., Koonin E.V., Smith D.R.;  
 RT "Genome sequence and comparative analysis of the solvent-producing  
 bacterium Clostridium acetobutylicum."  
 RL J. Bacteriol. 183:4823-4838(2001).  
 DR EMBL: AE007595; AAK78775.1; -.  
 DR InterPro: IPR003817; PS\_Dcarboxylase.  
 DR Pfam: PF02666; PS\_Dcarboxylase; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 291 AA; 34070 MW; 6CC2F1BCC87489D CRC64;

Query Match 87.9%; Score 29; DB 16; Length 291;  
 Best Local Similarity 71.4%; Pred. No. 78;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSK 7  
 :|||||  
 DB 43 TRIFSK 49

RESULT 7  
 09BSH4 PRELIMINARY; PRT; 297 AA.  
 AC 09BSH4;  
 DT 01-JUN-2001 (TEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
 DE SIMILAR TO RIKEN CDNA 2310066118 GENE (UNKNOWN) (PROTEIN FOR  
 MGC:13286).  
 DE Homo sapiens (human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE-PLACENTA, CHORIOCARCINOMA;  
 RC Straussberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN, NEUROBLASTOMA;  
 RA Straussberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC005049; AAH05049.1; -.  
 DR EMBL: BC007744; AAH07744.1; -.  
 DR InterPro: IPR002876; DUF28.  
 DR Pfam: PF01709; DUF28; 1.  
 DR ProDom: PD004323; DUF28; 1.  
 SQ SEQUENCE 297 AA; 32477 MW; 079C4716F32E6FF CRC64;

Query Match 87.9%; Score 29; DB 4; Length 297;  
 Best Local Similarity 85.7%; Pred. No. 79;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSK 7  
 :|||||  
 DB 78 SRIFSKL 84

RESULT 8  
 017134 PRELIMINARY; PRT; 352 AA.  
 ID 017134

```

AC 017134;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHEMETICAL 40.6 KDA PROTEIN.
GN F31F4.14.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
  investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Blanchard M., Kramer J., Elliott G., Twyman B.;
RT "The sequence of C. elegans cosmid F31F4.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Materon R.;
RT "Direct Submision.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF024503; AAG24092.1; -.
DR InterPro: IPR003839; DUF215.
DR Pfam: PF02688; DUF215; 1.
DR Hypothetical protein.
SQ SEQUENCE 352 AA; 40599 MW; BADAA1293C9B7585 CRC64;

Query Match
Best Local Similarity 87.9%; Score 29; DB 5; Length 352;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRIFSKL 7
Db 48 SRIFSKL 54

RESULT 9
Q9BMX4 PRELIMINARY; PRT; 419 AA.
AC Q9BMX4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE URIDINE KINASE/URACIL PHOSPHORIBOSYLTRANSFERASE.
GN UKUPRT.
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriida;
OC Cryptosporididae; Cryptosporidium.
OX NCBI_Taxid=5807;
RN [1]
RP SEQUENCE FROM N.A.
RA Woods K.M., Upton S.J.;
RT "Sequence of the uridine kinase/uracil phosphoribosyltransferase Gene
  from Cryptosporidium parvum.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF298155; AAG35652.1; -.
DR HSSP: Q26998; 1BD3.
DR InterPro: IPR001324; PRK.
DR InterPro: IPR000764; Uridine_kin.
DR Pfam: PF00485; PRK; 1.
DR PRINTS: PRO0988; URIDINKINASE.
KM Transferase: Kinase: Glycosyltransferase.
SQ SEQUENCE 419 AA; 47640 MW; 321BC8B143B545CF CRC64;

```

```

Query Match
Best Local Similarity 87.9%; Score 29; DB 5; Length 419;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRIFSKL 7
Db 47 TRIFSEL 53

RESULT 10
Q93UZO PRELIMINARY; PRT; 457 AA.
ID Q93UZO;
AC Q93UZO;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHEMETICAL 49.6 KDA PROTEIN.
OS Chromatium vinosum.
OC Bacteria; Proteobacteria; gamma subdivision; Chromatiaceae;
OC Allochromatium.
OX NCBI_Taxid=1049;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D;
RA Nagashima S., Shimada K., Matsuura K., Nagashima K.V.;
RT "Transcription of triplicated genes for the core light-harvesting
  complexes in the purple bacterium Allochromatium vinosum puf operon.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-D;
RA Corson G.E., Nagashima K.V., Matsuura K., Sakuragi Y., Wettasinghe R.,
  Qin H., Allen R., Knaff D.B.;
RT "Genes encoding light-harvesting and reaction center proteins from
  Chromatium vinosum.";
RL Photosyn. Res. 59:39-52(1999).
DR EMBL: AB050620; BAB44150.1; -.
DR Hypothetical protein.
SQ SEQUENCE 457 AA; 49562 MW; 42CD47B503185D38 CRC64;

Query Match
Best Local Similarity 87.9%; Score 29; DB 2; Length 457;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRIFSKL 7
Db 305 TRVFTKL 311

RESULT 11
Q9FN09 PRELIMINARY; PRT; 579 AA.
ID Q9FN09;
AC Q9FN09;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PHOTORECEPTOR-INTERACTING PROTEIN-LIKE.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
  Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
  Sequence features of the regions of 1,191,918 bp covered by seventeen

```

RT Physically assigned PI clones."  
RL DNA Res. 4:401-414(1997).  
DR EMBL: AB007645; BAB09029.1;  
DR InterPro: IPR000210; BTB\_POL.  
DR InterPro: IPR004249; NPH3.  
DR Pfam: PF03000; NPH3.1.  
DR SMART: SM00225; BTB; 1.  
DR PROSITE: PSS0097; BTB; 1.  
KW Receptor.  
SQ SEQUENCE 579 AA; 64765 MW; FD24DB00C95497B1 CRC64;  
  
Query Match 87.9%; Score 29; DB 10; Length 579;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TRIPSKL 7  
DB 531 SRIPSKL 537  
  
RESULT 12  
ID Q93Y14 PRELIMINARY; PRT; 579 AA.  
AC Q93Y14;  
DT 01-DEC-2001 (TREMblrel. 19, Created)  
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE PHOTORECEPTOR-INTERACTING PROTEIN-LIKE.  
GN A15G67440 OR K8K14.18.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
Palum C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
Chen K., Chung M.K., Hayashizaki Y., Ishida J., Kamaya A., Kawai J.,  
Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinzaki K.,  
Ecker J., Theologis A., Davis R.W.;  
RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL: AY054694; AAK96885.1;  
RW Receptor.  
KW SEQUENCE 579 AA; 64764 MW; FBEDFA8DAADC3E93 CRC64;  
SQ  
  
Query Match 87.9%; Score 29; DB 10; Length 579;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TRIPSKL 7  
DB 531 SRIPSKL 537  
  
RESULT 13  
ID Q98RV2 PRELIMINARY; PRT; 590 AA.  
AC Q98RV2;  
DT 01-OCT-2001 (TREMblrel. 18, Created)  
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)  
DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)  
DE HYPOTHETICAL 71.5 KDA PROTEIN ORF590 FROM CHROMOSOME 1.  
GN ORF590.  
OS Guillardia theta (Cryptomonas ph.).  
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.  
OX NCBI\_TaxID=55529;  
RN [1]  
RP SEQUENCE FROM N.A.  
RL MEDLINE=21223349; PubMed=11323671;  
RX

RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T.,  
RA Wu X., Reith M., Cavalier-Smith T., Møller U.G.;  
RT "The highly reduced genome of an enslaved algal nucleus."  
RL Nature 410:1091-1096(2001).  
DR EMBL: AF165818; AAK39848.1;  
KW Hypothetical protein.  
SQ SEQUENCE 590 AA; 71520 MW; 4FDC746B1A7330FD CRC64;  
  
Query Match 87.9%; Score 29; DB 10; Length 590;  
Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TRIPSKL 7  
DB 203 SRIPSKL 209  
  
RESULT 14  
ID Q961B6 PRELIMINARY; PRT; 606 AA.  
AC Q961B6;  
DT 01-DEC-2001 (TREMblrel. 19, Created)  
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE D24529P.  
GN TAF60.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=I, CN BW SP;  
RA Stapleton M., Brokstein P., Hong L., Agdayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuno J., Paclet J., Parages V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Gelinkner S.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY051702; AAK93126.1;  
SQ SEQUENCE 606 AA; 65654 MW; 9D93C3496A23C0A3 CRC64;  
  
Query Match 87.9%; Score 29; DB 5; Length 606;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TRIPSK 6  
DB 351 TRIPSK 356  
  
RESULT 15  
ID Q17405 PRELIMINARY; PRT; 1071 AA.  
AC Q17405;  
DT 01-JAN-1999 (TREMblrel. 09, Created)  
DT 01-JAN-1999 (TREMblrel. 09, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE HYPOTHETICAL 120.6 KDA PROTEIN AC3.5 IN CHROMOSOME V.  
GN AC3.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderiinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA McMurtry A.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: WEAK, TO PEPTIDASE FAMILY M1.

DR EMBL: 271177; CAA94872.1; -.  
 DR MEROPS; M01.UNM; -.  
 DR WormPep; AC3.5; CE05135.  
 DR InterPro; IPR001930; Aladiptase.  
 DR InterPro; IPR000209; Peptidase\_S8.  
 DR Pfam: PF01433; Peptidase\_M1; 1.  
 DR PROSITE: PS00136; SUBTYLASE\_ASP; UNKNOWN\_1.  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 81 101 POTENTIAL.  
 SO SEQUENCE 1071 AA: 120605 MW: 09733f0A0ECF231F CRC64;

Query Match 87.9%; Score 29; DB 5; Length 1071;  
 Best Local Similarity 85.7%; Pred. No. 2.8e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TRIPSKL 7  
 |||||  
 Db 788 TRIPKRL 794

Search completed: July 15, 2002, 13:22:40  
 Job time: 1476 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 15, 2002, 12:57:54 ; Search time 228.39 seconds  
(without alignments)  
3.404 Million cell updates/sec

Title: US-09-712-819A-2  
Perfect score: 33  
Sequence: 1 TRIFSXL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: A\_Geneseq\_032802:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	29	87.9	297	22	AAU27642
2	29	87.9	297	22	AAU27642
3	29	87.9	303	22	AAU41320
4	29	87.9	591	15	AAU56489
5	29	87.9	591	17	AAU60800
6	29	87.9	591	18	AAU25026
7	29	87.9	592	22	AAU71746
8	28	84.8	126	22	AAU06982
9	28	84.8	170	21	AAU61781
10	28	84.8	186	20	AAU19628
11	28	84.8	207	22	AAU90104

12	28	84.8	381	16	AAU78271	Human h1pocampal
13	28	84.8	381	16	AAU78272	Rat hippocampal ne
14	28	84.8	381	16	AAU78273	Rat hippocampal ne
15	28	84.8	381	19	AAU41710	Rhesus monkey neur
16	28	84.8	381	21	AAU56888	Human Y2 receptor
17	28	84.8	381	22	AAU07921	Human neurotrophide
18	28	84.8	381	22	AAU07957	Human neurotrophide
19	28	84.8	381	22	AAU06689	Human neurotrophide
20	28	84.8	381	22	AAU85109	Human neurotrophide
21	28	84.8	381	22	AAU85110	Human neurotrophide
22	28	84.8	438	22	AAU81858	Drosophila melanog
23	28	84.8	438	22	AAU81859	Drosophila melanog
24	28	84.8	438	22	AAU81860	Drosophila wild-type
25	28	84.8	438	22	AAU81861	Drosophila wild-type
26	28	84.8	438	22	AAU81862	Drosophila mutant
27	28	84.8	438	22	AAU81863	Drosophila mutant
28	28	84.8	438	22	AAU81864	Drosophila mutant
29	28	84.8	438	22	AAU81865	Drosophila mutant
30	28	84.8	438	22	AAU81866	Drosophila mutant
31	28	84.8	438	22	AAU81867	Drosophila mutant
32	28	84.8	438	22	AAU81868	Drosophila mutant
33	28	84.8	604	22	AAU51655	Human SCP-1 mutan
34	28	84.8	976	22	AAU66581	Human SCP-1 mutan
35	28	84.8	1492	22	AAU31597	Human SCP-1 mutan
36	27	81.8	49	21	AAU59989	Human novel foetal
37	27	81.8	50	22	AAU21482	Human novel foetal
38	27	81.8	65	22	AAU02168	Human polypeptide
39	27	81.8	67	22	AAU03192	Human polypeptide
40	27	81.8	136	21	AAU24639	Human SDF encoded
41	27	81.8	137	21	AAU44092	Human cancer assoc
42	27	81.8	138	21	AAU24638	Human SDF encoded
43	27	81.8	140	21	AAU24637	Human SDF encoded
44	27	81.8	162	21	AAU43503	Human cancer assoc
45	27	81.8	181	21	AAU14002	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1	AAU27642	standard; Protein; 297 AA.
ID	AAU27642	
AC	AAU27642	
DP	18-DEC-2001	(first entry)
DE	Human protein AFP50993.	
XX	Human; AFP; cytosolic; Marfan's syndrome; thrombocytopaenia;	
XX	leukaemia; porphyria; Gilles de la Tourette's syndrome; immunogen;	
XX	squamous cell carcinoma; diabetes mellitus; Grave's disease;	
XX	colon cancer; Alzheimer's disease; epiphyseal dysplasia.	
XX	Homo sapiens.	
XX	WO20016748-A2.	
XX	13-SEP-2001.	
XX	05-MAR-2001; 2001MO-US07192.	
XX	03-MAR-2000; 2000US-187221P.	
XX	(ZYMO) ZYMOGENETICS INC.	
XX	Conklin DC, Presnell SR, Adler DA;	
XX	WPI; 2001-589943/66.	
XX	N-PSDB; AAS44543.	
XX	Novel AFP polypeptides and polynucleotides, useful for diagnostic and	
XX	therapeutic purposes, in cancer therapy and for screening modulator	
XX	PT	

PT compounds -  
XX  
PS Claim 2: Page 135-136; 220pp; English.  
XX  
CC The invention relates to novel human AFP proteins (not defined) and  
CC the nucleic acids that encode them. AFP proteins are useful as standards  
CC in assays of protein and protein inhibitors in both clinical and research  
CC settings, as protein and amino acid supplements, including hydrolysates.  
CC The nucleic acids are useful for radiation hybrid mapping. Secretary  
CC fusion proteins of AFP are useful in cancer therapy, for enhancing in  
CC vitro cytotoxicity, for enhancing in vivo killing of target tissues, and  
CC for targeted cells or tissue inhibition or ablation. Anti-AFP antibodies  
CC are useful for isolating target polypeptides by affinity purification, in  
CC diagnostic assays for determining circulating or localised levels of  
CC target polypeptides, for tissue typing, for cell sorting, for screening  
CC expression libraries, for generating anti-idiotypic antibodies, and as  
CC neutralising antibodies or as antagonists to block protein activity in  
CC vitro and in vivo. AFP proteins and nucleic acids may be used to  
CC diagnose or treat (e.g. by gene therapy) diseases associated with the  
CC malfunction of the AFP e.g. AFP160924 and Marfan's syndrome,  
CC AFP576853/AFP39158 and thrombocytopaenia, leukaemia, porphyria, Gilles De  
CC La Tourette's syndrome, AFP652829 and squamous cell carcinoma, diabetes  
CC mellitus, Grave's disease, AFP664311 and colon cancer, AFP253034 and  
CC Alzheimer's disease and AFP686580 and epiphyseal dysplasia. Many more  
CC examples of disease are given in the specification. The present  
CC sequence represents an AFP of the invention.  
XX  
SQ Sequence 297 AA:  
  
Query Match 87.9%; Score 29; DB 22; Length 297;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 TRIFSXL 7  
Db 78 srlfskl 84  
  
RESULT 2  
AAM39534  
ID AAM39534 standard; Protein; 297 AA.  
AC AAM39534;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 2679.  
XX  
KW Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PE 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX

PA (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI: 2001-442253/47.  
DR N-PSDB; AAI58690.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Example 4: SEQ ID NO 2679; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cyostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC activation/inhibit activity, chemotactic/chemokine activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 297 AA:  
  
Query Match 87.9%; Score 29; DB 22; Length 297;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 TRIFSXL 7  
Db 78 srlfskl 84  
  
RESULT 3  
AAM41320  
ID AAM41320 standard; Protein; 303 AA.  
AC AAM41320;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 6251.  
XX  
KW Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PE 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR



PR 14-SEP-2000: 2000US-0662191.  
 PR 19-OCT-2000: 2000US-0693036.  
 PR 29-NOV-2000: 2000US-0727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Dimañac RT;  
 XX  
 DR WPI, 2001-442253/47.  
 DR N-PSDB: AAI60476.  
 XX  
 XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 PS  
 PS Example 2; SEQ ID NO 6251; 10078pp; English.  
 CC  
 CC The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AA436642-AA442213) with neurotropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 CC  
 CC  
 SO Sequence 303 AA;

Query Match 87.9%; Score 29; DB 22; Length 303;  
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRIFSK 7  
 :|||||  
 DB 84 srtfsk1 90

RESULT 4  
 AAR56489  
 ID AAR56489 standard; Protein; 591 AA.  
 XX  
 AC AAR56489;  
 XX  
 DT 23-MAR-1995 (first entry)  
 XX  
 DE TATA-binding protein-associated factor dTAFLI60.  
 XX  
 XX TATA-binding protein associated factor; dTAFLI60; screening;  
 KM diagnostic; therapeutic; gene transcription regulation.  
 KM  
 XX Drosophila.  
 OS  
 XX  
 XX WO9417087-A.  
 PN  
 XX  
 PD 04-AUG-1994.  
 XX  
 XX 28-JAN-1994; 94WO-US01114.  
 PF  
 XX 28-JAN-1993; 93US-0013412.  
 PR 30-JUN-1993; 93US-0087119.  
 XX  
 XX (REGC ) UNIV CALIFORNIA.  
 PA  
 XX

PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;  
 PI Tjian R, Wang E, Weinzierl ROJ;  
 XX  
 XX WPI, 1994-264019/32.  
 DR  
 DR N-PSDB: AAQ70727.  
 XX  
 XX TATA-binding protein associated protein factors - and  
 PT corresponding nucleotide sequence and deriv. antibodies, useful  
 PT in screening, diagnostics and therapeutics  
 XX  
 PS Disclosure; Page 85-88; 180pp; English.  
 CC  
 CC The TATA-binding protein associated factor dTAFLI60 (including  
 CC specific antibodies and fusion products) are used in drug screening,  
 CC diagnostics and therapeutics. They are used in the development of  
 CC specific biochemical assays for screening compounds that agonise or  
 CC antagonise selected transcription factors involved in regulating  
 CC gene expression associated with human pathology.  
 CC  
 CC  
 SO Sequence 591 AA;

Query Match 87.9%; Score 29; DB 15; Length 591;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRIFSK 6  
 :|||||  
 DB 337 trlfsk 342

RESULT 5  
 AAM06080  
 ID AAM06080 standard; Protein; 591 AA.  
 XX  
 AC AAM06080;  
 XX  
 DT 27-JAN-1997 (first entry)  
 XX  
 DE Drosophila TATA-binding protein associated factor dTAFLI60 protein.  
 XX  
 DE Drosophila; TATA-binding protein; TBP associated factor; TFIID;  
 KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;  
 KW holoenzyme; lambda-gt11; expression library.  
 XX  
 XX Drosophila melanogaster.  
 OS  
 XX  
 XX US534410-A.  
 PN  
 XX  
 PD 09-JUL-1996.  
 XX  
 XX 28-JAN-1993; 93US-0013412.  
 PF  
 XX 28-JAN-1994; 94US-0188582.  
 PR 28-JAN-1993; 93US-0013412.  
 PR 30-JUN-1993; 93US-0087119.  
 XX  
 XX (REGC ) UNIV CALIFORNIA.  
 PA  
 XX Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;  
 PI Tjian R, Wang E, Weinzierl ROJ;  
 XX  
 XX WPI, 1996-333245/33.  
 DR N-PSDB: AAT42213.  
 DR  
 XX  
 XX Screen for cpds. that bind human TATA-binding protein associated  
 PT factor - by testing ability to bind to polypeptide fragments of the  
 PT factor, useful as (ant)agonists of transcription factors involved in  
 PT disease.  
 XX  
 XX Examples; Column 53-58; 86pp; English.  
 PS  
 PS  
 CC This is the amino acid sequence encoding the Drosophila TATA-binding

CC Including transcription initiation. TAFs are nuclear proteins involved

592 AA; Sequence

Query Match	87.9%	Score 29;	DB 22;	Length 592;					
Best Local Similarity	100.0%;	Pred. No. 2.5e+02;							
Matches	6;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0

OY 1 TRIRSK 6  
| | | | |  
Db 337 trirfsk 342

## RESULT 8

AA006982  
ID AA006982 standard; Protein; 128 AA.

AC AA006982;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 20874.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

PN W0200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-0504927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HXYE-) HYSEQ INC.

P1 Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

DR N-PSDB; AAI86913.

XX Isolated nucleic acids and polypeptides, useful for preventing

PT diagnosing and treating e.g. leukaemia, inflammation and immune

PT disorders -

PS Claim 20; SEQ ID NO 20874; 1399PP + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 128 AA;

XX Query Match 84.8%; Score 28; DB 22; Length 128;

XX Best Local Similarity 83.3%; Pred. No. 85;

XX Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIRSK 6  
| | | | |  
Db 15 trirfsk 20

RESULT 9

AAG61781  
ID AAG61781 standard; Protein; 170 AA.

AC AAG61781;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 80243.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 22-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 04-MAY-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 07-MAY-1999; 99US-0132486.

PR 11-MAY-1999; 99US-0132487.

PR 14-MAY-1999; 99US-0132488.

PR 14-MAY-1999; 99US-0132489.

PR 18-MAY-1999; 99US-0132490.

PR 20-MAY-1999; 99US-0132491.

PR 21-MAY-1999; 99US-0132492.

PR 24-MAY-1999; 99US-0132493.

PR 25-MAY-1999; 99US-0132494.

PR 27-MAY-1999; 99US-0132495.

PR 28-MAY-1999; 99US-0132496.

PR 01-JUN-1999; 99US-0132497.

PR 03-JUN-1999; 99US-0132498.

PR 04-JUN-1999; 99US-0132499.

PR 07-JUN-1999; 99US-0132500.

PR 08-JUN-1999; 99US-0132501.

PR 10-JUN-1999; 99US-0132502.

PR 10-JUN-1999; 99US-0132503.

PR 14-JUN-1999; 99US-0132504.

PR 16-JUN-1999; 99US-0132505.

PR 16-JUN-1999; 99US-0132506.

PR 17-JUN-1999; 99US-0132507.

PR 18-JUN-1999; 99US-0132508.

PR 18-JUN-1999; 99US-0132509.

PR 18-JUN-1999; 99US-0132510.

PR 18-JUN-1999; 99US-0132511.

PR 18-JUN-1999; 99US-0132512.

PR 18-JUN-1999; 99US-0132513.

PR 18-JUN-1999; 99US-0132514.

PR 18-JUN-1999; 99US-0132515.

PR 18-JUN-1999; 99US-0132516.

PR 18-JUN-1999; 99US-0132517.

PR 18-JUN-1999; 99US-0132518.

PR 18-JUN-1999; 99US-0132519.

PR 18-JUN-1999; 99US-0132520.

PR 18-JUN-1999; 99US-0132521.

PR 18-JUN-1999; 99US-0132522.

PR 18-JUN-1999; 99US-0132523.

PR 18-JUN-1999; 99US-0132524.

PR 18-JUN-1999; 99US-0132525.

PR 18-JUN-1999; 99US-0132526.

PR 18-JUN-1999; 99US-0132527.

PR 18-JUN-1999; 99US-0132528.

PR 18-JUN-1999; 99US-0132529.

PR 18-JUN-1999; 99US-0132530.

PR 18-JUN-1999; 99US-0132531.

PR 18-JUN-1999; 99US-0132532.

PR 18-JUN-1999; 99US-0132533.

PR 18-JUN-1999; 99US-0132534.

PR 18-JUN-1999; 99US-0132535.

PR 18-JUN-1999; 99US-0132536.

PR 18-JUN-1999; 99US-0132537.

PR 18-JUN-1999; 99US-0132538.

PR 18-JUN-1999; 99US-0132539.

PR 18-JUN-1999; 99US-0132540.

PR 18-JUN-1999; 99US-0132541.

PR 18-JUN-1999; 99US-0132542.

PR 18-JUN-1999; 99US-0132543.

PR 18-JUN-1999; 99US-0132544.

PR 18-JUN-1999; 99US-0132545.

PR 18-JUN-1999; 99US-0132546.

PR 18-JUN-1999; 99US-0132547.

PR 18-JUN-1999; 99US-0132548.

PR 18-JUN-1999; 99US-0132549.

PR 18-JUN-1999; 99US-0132550.

PR 18-JUN-1999; 99US-0132551.

PR 18-JUN-1999; 99US-0132552.

PR 18-JUN-1999; 99US-0132553.

PR 18-JUN-1999; 99US-0132554.

PR 18-JUN-1999; 99US-0132555.

PR 18-JUN-1999; 99US-0132556.

PR 18-JUN-1999; 99US-0132557.

PR 18-JUN-1999; 99US-0132558.

PR 18-JUN-1999; 99US-0132559.

PR 18-JUN-1999; 99US-0132560.

PR 18-JUN-1999; 99US-0132561.

PR 18-JUN-1999; 99US-0132562.

PR 18-JUN-1999; 99US-0132563.

PR 18-JUN-1999; 99US-0132564.

PR 18-JUN-1999; 99US-0132565.

PR 18-JUN-1999; 99US-0132566.

PR 18-JUN-1999; 99US-0132567.

PR 18-JUN-1999; 99US-0132568.

PR 18-JUN-1999; 99US-0132569.

PR 18-JUN-1999; 99US-0132570.

PR 18-JUN-1999; 99US-0132571.

PR 18-JUN-1999; 99US-0132572.

PR 18-JUN-1999; 99US-0132573.

PR 18-JUN-1999; 99US-0132574.

PR 18-JUN-1999; 99US-0132575.

PR 18-JUN-1999; 99US-0132576.

PR 18-JUN-1999; 99US-0132577.

PR 18-JUN-1999; 99US-0132578.

PR 18-JUN-1999; 99US-0132579.

PR 18-JUN-1999; 99US-0132580.

PR 18-JUN-1999; 99US-0132581.

PR 18-JUN-1999; 99US-0132582.

PR 18-JUN-1999; 99US-0132583.

PR 18-JUN-1999; 99US-0132584.

PR 18-JUN-1999; 99US-0132585.

PR 18-JUN-1999; 99US-0132586.

PR 18-JUN-1999; 99US-0132587.

PR 18-JUN-1999; 99US-0132588.

PR 18-JUN-1999; 99US-0132589.

PR 18-JUN-1999; 99US-0132590.

PR 18-JUN-1999; 99US-0132591.

PR 18-JUN-1999; 99US-0132592.

PR 18-JUN-1999; 99US-0132593.

PR 18-JUN-1999; 99US-0132594.

PR 18-JUN-1999; 99US-0132595.

PR 18-JUN-1999; 99US-0132596.

PR 18-JUN-1999; 99US-0132597.

PR 18-JUN-1999; 99US-0132598.

PR 18-JUN-1999; 99US-0132599.

PR 18-JUN-1999; 99US-0132600.

PR 18-JUN-1999; 99US-0132601.

PR 18-JUN-1999; 99US-0132602.

PR 18-JUN-1999; 99US-0132603.

PR 18-JUN-1999; 99US-0132604.

PR 18-JUN-1999; 99US-0132605.

PR 18-JUN-1999; 99US-0132606.

PR 18-JUN-1999; 99US-0132607.

PR 18-JUN-1999; 99US-0132608.

PR 18-JUN-1999; 99US-0132609.

PR 18-JUN-1999; 99US-0132610.

PR 18-JUN-1999; 99US-0132611.

PR 18-JUN-1999; 99US-0132612.

PR 18-JUN-1999; 99US-0132613.

PR 18-JUN-1999; 99US-0132614.

PR 18-JUN-1999; 99US-0132615.

PR 18-JUN-1999; 99US-0132616.

PR 18-JUN-1999; 99US-0132617.

PR 18-JUN-1999; 99US-0132618.

PR 18-JUN-1999; 99US-0132619.

PR 18-JUN-1999; 99US-0132620.

PR 18-JUN-1999; 99US-0132621.

PR 18-JUN-1999; 99US-0132622.

PR 18-JUN-1999; 99US-0132623.

PR 18-JUN-1999; 99US-0132624.

PR 18-JUN-1999; 99US-0132625.

PR 18-JUN-1999; 99US-0132626.

PR 18-JUN-1999; 99US-0132627.

PR 18-JUN-1999; 99US-0132628.

PR 18-JUN-1999; 99US-0132629.

PR 18-JUN-1999; 99US-0132630.

PR 18-JUN-1999; 99US-0132631.

PR 18-JUN-1999; 99US-0132632.

PR 18-JUN-1999; 99US-0132633.

PR 18-JUN-1999; 99US-0132634.

PR 18-JUN-1999; 99US-0132635.

PR 18-JUN-1999; 99US-0132636.

PR 18-JUN-1999; 99US-0132637.

PR 18-JUN-1999; 99US-0132638.

PR 18-JUN-1999; 99US-0132639.

PR 18-JUN-1999; 99US-0132640.

PR 18-JUN-1999; 99US-0132641.

PR 18-JUN-1999; 99US-0132642.

PR 18-JUN-1999; 99US-0132643.

PR 18-JUN-1999; 99US-0132644.

PR 18-JUN-1999; 99US-0132645.

PR 18-JUN-1999; 99US-0132646.

PR 18-JUN-1999; 99US-0132647.

PR 18-JUN-1999; 99US-0132648.

PR 18-JUN-1999; 99US-0132649.

PR 18-JUN-1999; 99US-0132650.

PR 18-JUN-1999; 99US-0132651.

PR 18-JUN-1999; 99US-0132652.

PR 18-JUN-

PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	02-JUL-1999;	99US-0142055.	PR	20-SEP-1999;	99US-0154779.
PR	06-JUL-1999;	99US-0142390.	PR	22-SEP-1999;	99US-0155132.
PR	08-JUL-1999;	99US-0142803.	PR	23-SEP-1999;	99US-0155139.
PR	09-JUL-1999;	99US-0142820.	PR	24-SEP-1999;	99US-0155486.
PR	12-JUL-1999;	99US-0142877.	PR	28-SEP-1999;	99US-0155659.
PR	13-JUL-1999;	99US-0143542.	PR	29-SEP-1999;	99US-0155656.
PR	14-JUL-1999;	99US-0143624.	PR	04-OCT-1999;	99US-0157117.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157753.
PR	16-JUL-1999;	99US-0144086.	PR	07-OCT-1999;	99US-0158029.
PR	19-JUL-1999;	99US-0144325.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144331.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144334.	PR	13-OCT-1999;	99US-0159294.
PR	20-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159295.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159329.
PR	20-JUL-1999;	99US-0144684.	PR	14-OCT-1999;	99US-0159330.
PR	21-JUL-1999;	99US-0144814.	PR	14-OCT-1999;	99US-0159331.
PR	21-JUL-1999;	99US-0145086.	PR	14-OCT-1999;	99US-0159637.
PR	21-JUL-1999;	99US-0145088.	PR	18-OCT-1999;	99US-0159638.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0159584.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160770.
PR	23-JUL-1999;	99US-0145145.	PR	21-OCT-1999;	99US-0160814.
PR	23-JUL-1999;	99US-0145218.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160980.
PR	26-JUL-1999;	99US-0145276.	PR	22-OCT-1999;	99US-0160981.
PR	27-JUL-1999;	99US-0145913.	PR	22-OCT-1999;	99US-0160989.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161404.
PR	28-JUL-1999;	99US-0145951.	PR	25-OCT-1999;	99US-0161405.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161406.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161359.
PR	03-AUG-1999;	99US-0146389.	PR	26-OCT-1999;	99US-0161360.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161361.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161920.
PR	04-AUG-1999;	99US-0147302.	PR	28-OCT-1999;	99US-0161992.
PR	05-AUG-1999;	99US-0147192.	PR	29-OCT-1999;	99US-0162142.
PR	06-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148177.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			

---

PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	30-AUG-1999;	99US-0151080.
PR	31-AUG-1999;	99US-0151303.
PR	01-SEP-1999;	99US-0151438.
PR	07-SEP-1999;	99US-0151930.
PR	10-SEP-1999;	99US-0152363.
PR	13-SEP-1999;	99US-0153070.
PR	15-SEP-1999;	99US-0153758.
PR	16-SEP-1999;	99US-0154018.
PR	20-SEP-1999;	99US-0154039.
PR	22-SEP-1999;	99US-0154779.
PR	23-SEP-1999;	99US-0155132.
PR	24-SEP-1999;	99US-0155139.
PR	28-SEP-1999;	99US-0155486.
PR	29-SEP-1999;	99US-0155659.
PR	04-OCT-1999;	99US-0157117.
PR	06-OCT-1999;	99US-0157753.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	14-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	18-OCT-1999;	99US-0159638.
PR	21-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	26-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	28-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	29-OCT-1999;	99US-0162142.

Query Match

Best local Similarity 71.4%; Score 28; DB 21; Length 170;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY

1 TRRFSL 7

128 svtfsl 134

Db

128 svtfsl 134

RESULT 10

AA19828

AA19828 standard; Protein: 186 AA.

XX

AA19828:

XX

19-JUL-1999 (first entry)

DT

XX

DE	B.	burgdorferi antigenic protein, f82.a.
XX		
KW	Antigenic protein; vaccine; Lyme disease; infection; detection.	
XX		
OS	Borrelia burgdorferi <sup>1</sup> .	
PN		
XN	MO9859071-A1.	
PD		
PF	30-DEC-1998.	
XX		
PE	18-JUN-1998;	98WO-US12718.
XX		
PR	03-SEP-1997;	97US-0057483.
PR	20-JUN-1997;	97US-0050359.
PR	22-JUL-1997;	97US-0053344.
PR	22-JUL-1997;	97US-0053377.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
PA	(MEDI-) MEDIMUNE INC.	
XX		
PI	Choi GH, Erwin AL, Hanson MS, Iatigra R;	
XX		
DR	WPJ, 1999-189980/16.	
DR	N-PDSB; AAK61525.	
XX		
PT	New isolated Borrelia burgdorferi nucleic acids - used to develop	
PT	products for the diagnosis, prevention and treatment of diseases	
PT	caused by Borrelia, particularly Lyme disease	
XX		
PS	Claim 12; Page 82; 275pp; English.	
XX		
CC	This sequence represents a Borrelia burgdorferi (Bb) protein of the	
CC	invention, which is suitable for use in a vaccine. The Bb polypeptides	
CC	can be used in vaccines for eliciting protective antibodies to members of	
CC	the Borrelia genus, particularly for the use against Lyme disease in	
CC	humans and animals. They can be used for preventing or attenuating an	
CC	infection caused by a member of the Borrelia genus. The products can also	
CC	be used for detection of members of the Borrelia genus.	
XX		
SQ	Sequence	186 AA;

```

Query Match      84.8%; Score 28; DB 20; Length 186;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches      5; Conservative      1; Mismatches      0; Indels      0; Gaps      0;

QY      1 TRFSK 6
      ||:| | |
Db      2 trfsk 7

RESULT 11
AAG90104
ID      AAG90104 standard; Protein; 207 AA.
XX
AC      AAG90104;
XX
DP      26-SEP-2001 (first entry)
XX
DE      C glutamicum protein fragment SEQ ID NO: 3858.
XX
KW      Corynebform bacterium; amino acid synthesis; vitamin; saccharide;
KW      organic acid synthesis.
XX
OS      Corynebacterium glutamicum.
XX
PN      EP1108790-A2.
XX
PD      20-JUN-2001.
XX
PF      18-DEC-2000; 2000EP-0127688.
XX
PR      16-DEC-1999; 99JP-0377484.

```

07-APR-2000: 2000JP-0159162.  
PR 03-AUG-2000: 2000JP-0280988.  
XX  
XX (KYOW ) KYOMA HAKKO KOGYO KK.  
XX  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
XX WPI: 2001-376931/40.  
DR  
DR N-PSDB: AAH65323.  
XX  
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
PI mutation point of a gene, measuring expression of a gene, analysing  
PI expression profile or pattern of a gene and identifying homologous gene  
PT  
XX  
XX Claim 17: SEQ ID NO: 3858; 246bp + Sequence listing; English.  
XX  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of corynebacterium bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived  
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
XX Sequence 207 AA;

Query Match	84.88;	Score 28;	DB 22;	Length 207;
Best Local Similarity	83.38;	Pred. No. 1.4e+02;		
Matches	5;	Conservative	1;	Mismatches 0;
				Indels 0;
				Gaps 0;
Oy	1	TRIFSK 6		
Db	127	trvfsk 132		
RESULT 12				
AAr78271				
ID	AAr78271	standard; Protein; 381 AA.		
XX				
AC	AAr78271;			
XX				
DT	06-MAR-1996	(first entry)		
XX				
DE	Human hippocampal neuropeptide Y2 receptor (clone CG-13).			
XX				
KW	Hippocampus; brain; neuropeptide; peptide; hypertension; pain;			
KW	gastrointestinal disorder; diagnosis; sleeping disorder; epilepsy;			
KW	memory loss; diarrhoea; nasal congestion.			
XX				
OS	Homo sapiens.			
XX				
PN	W09521245-A1.			
XX				
PD	10-AUG-1995.			
XX				
PF	03-FEB-1995;	95WO-US01469.		
XX				
PR	03-FEB-1994;	94US-0192288.		
XX				
PA	(SYNA-) SYNAPTIC PHARM CORP.			
XX				
PI	Branchek T, Gerald C, Walker MW, Weinschank R;			
XX				
DR	WPI; 1995-283765/37.			

DR N-PSDB; AAQ95034.  
XX  
XX Human and rat  $\gamma 2$  receptor DNA and protein - useful in diagnosis and  
PT treatment of e.g. cognitive and gastrointestinal disorder(s),  
PT hypertension and pain  
XX  
XX Disclosure: Page 115-117; 193pp; English.  
XX  
XX This protein, which may be expressed recombinantly in NIH3T3 cells,  
CC designated N-hy2-5 (ATCC CRL-11825) cells, using plasmid pCEXV-hy2,  
CC (ATCC 75659), may be useful in the treatment of several  
CC pathophysiological conditions including cognitive, gastrointestinal  
CC or sleeping disorders, epilepsy, hypertension, memory loss, diarrhoea,  
CC nasal congestion or pain.  
XX  
XX Sequence 381 AA;  
SQ

Query Match 84.8%; Score 28; DB 16; Length 381;  
Best Local Similarity 85.7%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSKL 7  
|||:||||  
DB 240 trlowski 246

## RESULT 13

AA878272  
ID AAR78272 standard; Protein; 381 AA.  
XX  
XX AAR78272;  
XX  
XX 06-MAR-1996 (first entry)  
XX  
XX

DE Rat hippocampal neuropeptide  $\gamma 2$  receptor (clone R55a).  
XX

KW Hippocampus; brain; neuropeptide; peptide; hypertension; pain;  
KM gastrointestinal disorder; diagnosis; sleeping disorder; epilepsy;  
XX memory loss; diarrhoea; nasal congestion.  
XX

OS Rattus rattus.  
XX

PN MO9521245-A1.  
XX

PD 10-AUG-1995.  
XX

PS 03-FEB-1995; 95WO-US01469.  
XX

PR 03-FEB-1994; 94US-0192288.  
XX

PA (SYNA-) SYNAPTIC PHARM CORP.  
XX

PI Branchek T, Gerald C, Walker MW, Weinschank R;  
XX

DR WPI; 1995-283765/37.  
XX

DR N-PSDB; AAQ95035.  
XX

PT Human and rat  $\gamma 2$  receptor DNA and protein - useful in diagnosis and  
PT treatment of e.g. cognitive and gastrointestinal disorder(s),  
PT hypertension and pain  
XX  
XX

PS Disclosure: Page 119-120; 193pp; English.  
XX

CC This protein, which may be expressed recombinantly in NIH3T3 cells,  
CC designated N-hy2-5 (ATCC CRL-11825) cells, using plasmid pCEXV-ry2a  
CC (ATCC 97035), may be useful in the treatment of several  
CC pathophysiological conditions including cognitive, gastrointestinal  
CC or sleeping disorders, epilepsy, hypertension, memory loss, diarrhoea,  
CC nasal congestion or pain.  
XX  
XX

SQ Sequence 381 AA;  
DT

Query Match 84.8%; Score 28; DB 16; Length 381;  
Best Local Similarity 85.7%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSKL 7  
|||:||||  
DB 240 trlowski 246

## RESULT 14

AA878273  
ID AAR78273 standard; Protein; 381 AA.  
XX  
XX AAR78273;  
XX  
XX 06-MAR-1996 (first entry)  
XX  
XX

DE Rat hippocampal neuropeptide  $\gamma 2$  receptor (clone R526a).  
XX

KW Hippocampus; brain; neuropeptide; peptide; hypertension; pain;  
KM gastrointestinal disorder; diagnosis; sleeping disorder; epilepsy;  
XX memory loss; diarrhoea; nasal congestion.  
XX

OS Rattus rattus.  
XX

PN MO9521245-A1.  
XX

PD 10-AUG-1995.  
XX

PS 03-FEB-1995; 95WO-US01469.  
XX

PR 03-FEB-1994; 94US-0192288.  
XX

PA (SYNA-) SYNAPTIC PHARM CORP.  
XX

PI Branchek T, Gerald C, Walker MW, Weinschank R;  
XX

DR WPI; 1995-283765/37.  
XX

DR N-PSDB; AAQ95036.  
XX

PT Human and rat  $\gamma 2$  receptor DNA and protein - useful in diagnosis and  
PT treatment of e.g. cognitive and gastrointestinal disorder(s),  
PT hypertension and pain  
XX  
XX

PS Disclosure: Page 122-124; 193pp; English.  
XX

CC This protein, which may be expressed recombinantly in NIH3T3 cells,  
CC designated N-hy2-5 (ATCC CRL-11825) cells, using plasmid pCEXV-ry2b  
CC (ATCC 97036), may be useful in the treatment of several  
CC pathophysiological conditions including cognitive, gastrointestinal  
CC or sleeping disorders, epilepsy, hypertension, memory loss, diarrhoea,  
CC nasal congestion or pain.  
XX  
XX

SQ Sequence 381 AA;  
DT

Query Match 84.8%; Score 28; DB 16; Length 381;  
Best Local Similarity 85.7%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSKL 7  
|||:||||  
DB 240 trlowski 246

## RESULT 15

AA841710  
ID AAW41710 standard; Protein; 381 AA.  
XX  
XX AAW41710;  
XX

DT 22-JUN-1998 (first entry)

XX Rhesus monkey neuropeptide Y2 receptor.  
 XX  
 XX Neuropeptide Y2 receptor; rhesus monkey; neurotransmitter; angina;  
 KW pain; renal failure; neurodegeneration; metabolic disorder;  
 KW obesity; inflammation; asthma; vasospasm; heart failure; shock;  
 KW cardiac hypertrophy; hypertension; myocardial infarction;  
 KW arrhythmia; epilepsy; cerebral infarction; stroke; hemorrhage;  
 KW cerebral vasospasm; depression; anxiety; schizophrenia; dementia;  
 KW incontinence; Crohn's disease; anorexia; bronchoconstriction;  
 KW therapy.  
 XX  
 XX Macaca mulatta.  
 XX  
 XX MO9748406-A1.  
 XX  
 XX 24-DEC-1997.  
 XX  
 XX 17-JUN-1997; 97MO-US10048.  
 XX  
 XX 17-JUN-1996; 96US-0019809.  
 XX  
 XX (ELIL ) LILLY & CO ELI.  
 XX  
 XX Baez M, Yang P;  
 XX  
 XX WPI; 1998-062842/06.  
 XX  
 XX N-PSDB; AAV05330-31.  
 XX  
 XX Rhesus monkey neuropeptide Y2 receptor - useful to develop  
 PT treatments for, e.g. angina, renal failure, neurodegeneration, pain,  
 PT obesity, metabolic disorder, inflammation or asthma  
 XX  
 XX Claim 1: Page 49-50; 57pp; English.  
 XX  
 XX This protein comprises a novel rhesus monkey receptor, designated  
 CC neuropeptide Y2 receptor, that has affinity for neuropeptide Y,  
 CC pancreatic polypeptide and peptide YY. Y2 receptors are found in  
 CC a variety of brain regions as well as in the peripheral nervous  
 CC system. In all tissues, Y2 receptors mediate a decrease in the  
 CC release of neurotransmitters. Also claimed are cDNA and RNA  
 CC sequences (see AAV05330-31), expression vectors, host cells  
 CC (especially Escherichia coli or mammalian cells), and methods for  
 CC evaluating the effectiveness of test compounds for the treatment or  
 CC prevention of conditions associated with an excess or deficiency of  
 CC stimulation of the receptor. Such compounds can be used to treat or  
 CC prevent disorders or diseases pertaining to the heart, blood vessels  
 CC or the renal system (e.g. vasospasm, heart failure, shock, cardiac  
 CC hypertrophy, increased blood pressure, angina, myocardial  
 CC infarction, sudden cardiac death, arrhythmia, peripheral vascular  
 CC disease and abnormal renal conditions such as impaired flow of  
 CC fluid, abnormal mass transport or renal failure), conditions  
 CC related to increased sympathetic nerve activity (e.g. during or  
 CC after coronary artery surgery and operations and surgery in the  
 CC gastrointestinal tract), cerebral diseases and diseases related to  
 CC the central nervous system (e.g. epilepsy, cerebral infarction,  
 CC neurodegeneration, stroke, cerebral vasospasm and hemorrhage,  
 CC depression, anxiety, schizophrenia and dementia), conditions related  
 CC to pain or nociception, diseases related to abnormal  
 CC gastrointestinal motility and secretion (e.g. different forms of  
 CC ileus, urinary incontinence, and Crohn's disease, abnormal drink  
 CC and food intake disorders, such as obesity, anorexia, bulimia and  
 CC metabolic disorders), diseases related to sexual dysfunction and  
 CC reproductive disorders, conditions or disorders associated with  
 CC inflammation, respiratory diseases (e.g. bronchoconstriction and  
 CC asthma) and diseases related to abnormal hormone release (e.g.  
 CC luteinizing hormone, growth hormone, insulin and prolactin).  
 XX  
 XX Sequence 381 AA:  
 SQ

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 TRFSKL 7  
 |||:|  
 Db 240 trlwskl 246

Search completed: July 15, 2002, 12:57:55  
 Job time: 411 sec

Query Match 84.8%; Score 28; DB 19; Length 381;  
 Best Local Similarity 85.7%; Pred. No. 2.6e+02;





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 12:59:16 ; Search time 75.67 Seconds  
(without alignments)  
2.260 Million cell updates/sec

Title: US-09-712-819A-2

Perfect score: 33

Sequence: 1 TRIPSKL 7

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2.6/prodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2.6/prodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2.6/prodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2.6/prodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2.6/prodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2.6/prodata/2/1aa/Backfilltest.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	29	87.9	591	1	US-08-188-582-7	Sequence 7, Appl1
2	29	87.9	591	1	US-08-646-715-7	Sequence 7, Appl1
3	28	84.8	381	1	US-08-192-288-2	Sequence 2, Appl1
4	28	84.8	381	2	US-08-687-355A-2	Sequence 2, Appl1
5	28	84.8	381	2	US-08-687-355A-4	Sequence 4, Appl1
6	28	84.8	381	2	US-08-687-355A-6	Sequence 6, Appl1
7	28	84.8	381	4	US-09-200-673-16	Sequence 16, Appl1
8	28	84.8	381	4	US-08-876-798A-2	Sequence 2, Appl1
9	28	84.8	976	4	US-09-104-324B-4	Sequence 4, Appl1
10	27	81.8	584	1	US-08-313-288B-17	Sequence 17, Appl1
11	26	78.8	217	4	US-08-297-431B-33	Sequence 33, Appl1
12	26	78.8	411	2	US-08-440-845D-3	Sequence 3, Appl1
13	26	78.8	411	4	US-08-868-458-3	Sequence 3, Appl1
14	26	78.8	411	4	US-09-522-800-16	Sequence 16, Appl1
15	26	78.8	476	4	US-08-924-183-1	Sequence 1, Appl1
16	26	78.8	476	4	US-08-924-183-2	Sequence 2, Appl1
17	26	78.8	476	4	US-08-924-183-7	Sequence 7, Appl1
18	26	78.8	476	4	US-09-488-364-1	Sequence 1, Appl1
19	26	78.8	476	4	US-09-488-364-2	Sequence 2, Appl1
20	26	78.8	476	4	US-09-488-364-7	Sequence 7, Appl1
21	26	78.8	479	3	US-08-278-635B-2	Sequence 2, Appl1
22	26	78.8	479	3	US-08-464-258B-2	Sequence 2, Appl1
23	26	78.8	479	3	US-08-471-961-2	Sequence 2, Appl1
24	26	78.8	484	4	US-08-913-578-2	Sequence 2, Appl1
25	26	78.8	484	4	US-08-785-427-2	Sequence 2, Appl1
26	26	78.8	729	2	US-08-677-298-2	Sequence 2, Appl1
27	26	78.8	946	3	US-08-560-005-4	Sequence 4, Appl1

28	26	78.8	946	4	US-09-418-540-4	Sequence 4, Appl1
29	25	75.8	112	4	US-09-404-671-8	Sequence 8, Appl1
30	25	75.8	181	1	US-08-418-444A-3	Sequence 3, Appl1
31	25	75.8	218	4	US-09-651-656-105	Sequence 105, App
32	25	75.8	257	4	US-09-183-217-2	Sequence 2, Appl1
33	25	75.8	509	3	US-08-369-822C-26	Sequence 26, Appl1
34	25	75.8	509	3	US-08-582-776C-41	Sequence 41, Appl1
35	25	75.8	509	3	US-08-434-831B-38	Sequence 38, Appl1
36	24	72.7	213	1	US-08-287-442-2	Sequence 2, Appl1
37	24	72.7	213	1	US-08-459-701-2	Sequence 2, Appl1
38	24	72.7	213	1	US-08-450-298-2	Sequence 2, Appl1
39	24	72.7	213	1	US-08-459-174-2	Sequence 2, Appl1
40	24	72.7	213	1	US-08-761-258-3	Sequence 3, Appl1
41	24	72.7	213	1	US-08-761-258-5	Sequence 5, Appl1
42	24	72.7	213	1	US-08-761-258-10	Sequence 10, Appl1
43	24	72.7	213	2	US-09-009-218-2	Sequence 2, Appl1
44	24	72.7	213	2	US-09-009-218-4	Sequence 4, Appl1
45	24	72.7	213	2	US-09-009-218-6	Sequence 6, Appl1

## ALIGNMENTS

RESULT 1  
US-08-188-582-7  
Sequence 7, Application US/08188582  
Patent No. 5534410  
GENERAL INFORMATION:  
APPLICANT: Tjian, Robert  
APPLICANT: Comai, Lucio  
APPLICANT: Dynlacht, Brian D.  
APPLICANT: Hoey, Timothy  
APPLICANT: Ruppert, Siegfried  
APPLICANT: Tanese, Naoko  
APPLICANT: Wang, Edith  
APPLICANT: Weinzierl, Robert O.J.  
TITLE OF INVENTION: YATA-BINDING PROTEIN ASSOCIATED FACTORS,  
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/188,582  
FILING DATE: 28-JAN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RMO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 591 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-188-582-7  
Query Match 87.9% ; Score 29 ; DB 1 ; Length 591 ;

Best Local Similarity 100.0%; Pred. No. 51;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSK 6  
Db 337 TRIFSK 342

## RESULT 2

US-08-646-715-7  
; Sequence 7, Application US/08646715  
; Patent No. 5637686  
; GENERAL INFORMATION:  
; APPLICANT: Tjian, Robert  
; APPLICANT: Comai, Lucio  
; APPLICANT: Dynlacht, Brian D.  
; APPLICANT: Hoey, Timothy  
; APPLICANT: Ruppert, Siegfried  
; APPLICANT: Tanese, Naoko  
; APPLICANT: Wang, Edith  
; APPLICANT: Weinzierl, Robert O.J.  
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/646,715  
; FILING DATE: 09-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/188,582  
; FILING DATE: 28-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A  
; REGISTRATION NUMBER: 36,627  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 396-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 591 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-646-715-7

Query Match 87.9%; Score 29; DB 1; Length 591;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSK 6  
Db 337 TRIFSK 342

RESULT 3  
US-08-192-288-2  
; Sequence 2, Application US/08192288  
; Patent No. 5545549

; GENERAL INFORMATION:  
; APPLICANT: Gerald, Christophe  
; APPLICANT: Walker, Mary  
; APPLICANT: Branchek, Theresa  
; APPLICANT: Weinshank, Richard L.  
; TITLE OF INVENTION: DNA ENCODING A HUMAN NEUROPEPTIDE  
; TITLE OF INVENTION: Y/PEPTIDE YY (Y2) RECEPTOR AND USES THEREOF  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/192,288  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 44742/JPW/TEP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 664-0525  
; TELEX: 422523 COOP UT  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 381 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-192-288-2

Query Match 84.8%; Score 28; DB 1; Length 381;  
Best Local Similarity 85.7%; Pred. No. 56;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIFSKL 7  
Db 240 TRIFSKL 246

RESULT 4  
US-08-687-355A-2  
; Sequence 2, Application US/08687355A  
; Patent No. 5989834  
; GENERAL INFORMATION:  
; APPLICANT: Syntactic Pharmaceutical Corporation  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING NEUROPEPTIDE  
; TITLE OF INVENTION: Y/PEPTIDE YY (Y2) RECEPTORS AND USES THEREOF  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/687,355A  
FILING DATE: No. 5989834ember 26, 1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 44742-A-PCT/JPM/MAT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 381 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-687-355A-2

Query Match 84.8%; Score 28; DB 2; Length 381;  
Best Local Similarity 85.7%; Pred. No. 56;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIESKL 7  
|||||  
DB 240 TRIMSKL 246

RESULT 5  
US-08-687-355A-4  
Sequence 4, Application US/08687355A  
Patent No. 5989834  
GENERAL INFORMATION:  
APPLICANT: Synaptic Pharmaceutical Corporation  
TITLE OF INVENTION: NUCLEIC ACID ENCODING NEUROPEPTIDE  
NUMBER OF SEQUENCES: 27  
NUMBER OF INVENTIONS: Y/PEPTIDE YY (Y2) RECEPTORS AND USES THEREOF  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,355A  
FILING DATE: No. 5989834ember 26, 1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 44742-A-PCT/JPM/MAT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 381 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-687-355A-4

Query Match 84.8%; Score 28; DB 2; Length 381;  
Best Local Similarity 85.7%; Pred. No. 56;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIESKL 7  
|||||  
DB 240 TRIMSKL 246

RESULT 6  
US-08-687-355A-6  
Sequence 6, Application US/08687355A  
Patent No. 5989834  
GENERAL INFORMATION:  
APPLICANT: Synaptic Pharmaceutical Corporation  
TITLE OF INVENTION: NUCLEIC ACID ENCODING NEUROPEPTIDE  
NUMBER OF SEQUENCES: 27  
NUMBER OF INVENTIONS: Y/PEPTIDE YY (Y2) RECEPTORS AND USES THEREOF  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,355A  
FILING DATE: No. 5989834ember 26, 1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 44742-A-PCT/JPM/MAT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 381 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-687-355A-6

Query Match 84.8%; Score 28; DB 2; Length 381;  
Best Local Similarity 85.7%; Pred. No. 56;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TRIESKL 7  
|||||  
DB 240 TRIMSKL 246

RESULT 7  
US-09-200-673-16  
Sequence 16, Application US/09200673A  
Patent No. 6316203  
GENERAL INFORMATION:  
APPLICANT: Gerald, Christophe P.G.  
APPLICANT: Weinschenk, Richard L.  
APPLICANT: Walker, Mary W.  
APPLICANT: Branchek, Theresa  
TITLE OF INVENTION: Methods of Modifying Feeding Behavior, Compounds Useful  
TITLE OF INVENTION: In Such Methods, and DNA Encoding a Hypothalamic  
TITLE OF INVENTION: Atypical Neuropeptide Y/Peptide YY Receptor (Y5)  
FILE REFERENCE: 46166-B2/JPM  
CURRENT APPLICATION NUMBER: US/09/200,673A  
CURRENT FILING DATE: 1998-11-25  
EARLIER APPLICATION NUMBER: 08/566,096  
EARLIER FILING DATE: 1995-12-01  
EARLIER APPLICATION NUMBER: 08/349,025

EARLIER FILING DATE: 1994-12-02  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 16  
LENGTH: 381  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-200-673-16

Query Match 84.8%; Score 28; DB 4; Length 381;  
Best Local Similarity 85.7%; Pred. NO. 56;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRISKL 7  
1111111  
DB 240 TRIMSKL 246

RESULT 8  
US-08-876-798A-2  
Sequence 2, Application US/08876798A  
Patent No. 6353478  
GENERAL INFORMATION:  
APPLICANT: Baez, Melvyn  
APPLICANT: Yang, Pely1  
TITLE OF INVENTION: RHESUS MONKEY NEUROPEPTIDE Y RECEPTOR  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: U.S.A.  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/876,798A  
FILING DATE: 16-JUN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gaylo, Paul J.  
REGISTRATION NUMBER: 36,808  
REFERENCE/DOCKET NUMBER: X-10901  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317)276-0756  
TELEFAX: (317)276-2763  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 381 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-876-798A-2

Query Match 84.8%; Score 28; DB 4; Length 381;  
Best Local Similarity 85.7%; Pred. NO. 56;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRISKL 7  
1111111  
DB 240 TRIMSKL 246

RESULT 9  
US-09-104-324B-4  
Sequence 4, Application US/09104324B  
Patent No. 6232460

GENERAL INFORMATION:  
APPLICANT: T Tecl, Ozlem; Sahin, Ugur; Pfeundschuh, Michael  
TITLE OF INVENTION: Methods for Diagnosis And Treating Cancers,  
TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample  
TITLE OF INVENTION: No. 6232460mal Cells  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski LLP  
STREET: 666 Fifth Avenue  
CITY: New York City  
STATE: New York  
ZIP: 10103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Morpheus  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/104,324B  
FILING DATE: 25-June-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/892,702  
FILING DATE: 15-July-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. 6232460man D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LOD 5491  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 318-3000  
TELEFAX: (212) 752-5958  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 976 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-09-104-324B-4

Query Match 84.8%; Score 28; DB 4; Length 976;  
Best Local Similarity 71.4%; Pred. NO. 1,4e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRISKL 7  
1111111  
DB 103 SRVSKL 109

RESULT 10  
US-08-313-288B-17  
Sequence 17, Application US/08313288B  
Patent No. 5750502  
GENERAL INFORMATION:  
APPLICANT: Jessell, Thomas M. and Avihu Klar  
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPOINDIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/313,288B  
FILING DATE: January 5, 1995  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28, 678  
REFERENCE/DOCKET NUMBER: 40028-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
TELEX:  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 584 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-313-288B-17

Query Match 81.8%; Score 27; DB 1; Length 584;  
Best Local Similarity 71.4%; Pred. No. 1.4e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRISKL 7  
||||:|  
Db 290 TRIFTKV 296

RESULT 11  
US-08-297-431B-33  
Sequence 33, Application US/08297431B  
Patent No. 6136605  
GENERAL INFORMATION:  
APPLICANT: Fahl, William E  
APPLICANT: Gulick, Andrew M  
APPLICANT: Manoharan, T Herbert  
APPLICANT: Puchalski, Ralph B  
APPLICANT: Kramert, Katharine  
APPLICANT: Wasserman Wyeth M  
TITLE OF INVENTION: NOVEL GLUTATHIONE S-TRANSFERASE ISOFORMS  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saul, Ewing, Remick & Saul, LLP  
STREET: 1500 Market Street, 38th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19102-2186  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/297,431B  
FILING DATE: August 26, 1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Reed, Janet E.  
REGISTRATION NUMBER: 36252  
REFERENCE/DOCKET NUMBER: WARF F039  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-972-8386  
TELEFAX: 215-972-2292  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 217 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-297-431B-33

Query Match 78.8%; Score 26; DB 4; Length 217;

Best Local Similarity 85.7%; Pred. No. 93;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TRISKL 7  
|||||  
Db 205 TRIFSKL 211

RESULT 12  
US-08-440-845D-3  
Sequence 3, Application US/08440845D  
Patent No. 5953329  
GENERAL INFORMATION:  
APPLICANT: Yuan, L.  
APPLICANT: Kridl, J.  
APPLICANT: Debesb, K.  
APPLICANT: Knauf, V.  
TITLE OF INVENTION: Engineering Plant Thioesterases For  
TITLE OF INVENTION: Altered Substrate Specificity.  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Calgene, Inc.  
STREET: 1920 Fifth Street  
CITY: Davis  
STATE: CA  
COUNTRY: USA  
ZIP: 95616  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.40 MB  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 7.0  
SOFTWARE: Microsoft Word 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/440,845D  
FILING DATE: 15-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Donna E. Scherer  
REGISTRATION NUMBER: 34,719  
NAME: Carl J. Schwedler  
REGISTRATION NUMBER: 36,924  
REFERENCE/DOCKET NUMBER: CGNE 113  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (916) 753-6313  
TELEFAX: (916) 753-1510  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 411 residues  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-440-845D-3

Query Match 78.8%; Score 26; DB 2; Length 411;  
Best Local Similarity 85.7%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TRISKL 7  
|||||  
Db 259 TRIFSKL 265

RESULT 13  
US-08-868-458-3  
Sequence 3, Application US/08868458  
Patent No. 6150512  
GENERAL INFORMATION:  
APPLICANT: Yuan, L.  
TITLE OF INVENTION: Engineering Plant Thioesterases And  
Disclosure of Plant Thioesterases  
Having No. 6150512el Substrate Specificity  
NUMBER OF SEQUENCES: 19

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Calgene, Inc.
STREET: 1920 Fifth Street
CITY: Davis
STATE: CA
COUNTRY: USA
ZIP: 95616

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
COMPUTER: IBM PC
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: Microsoft Word For Window 95 7.0a

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/868,458
FILING DATE: 03-Jun-1997
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/07064
FILING DATE: 15-MAY-96
APPLICATION NUMBER: 08/537,083
FILING DATE: 29-SEPT-1995

ATTORNEY/AGENT INFORMATION:
NAME: Carl J. Schwedler
REGISTRATION NUMBER: 36,924
REFERENCE/DOCKET NUMBER: CGNE 133
TELECOMMUNICATION INFORMATION:
TELEPHONE: (916) 753-6313
TELEFAX: (916) 753-1510
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 411 residues
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-08-868-458-3

Query Match      78.8%; Score 26; DB 4; Length 411;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TRIFSKL 7
      || |||
Db      259 TRRSKL 265

RESULT 14
US-09-522-800-16
; Sequence 16, Application US/09522800
; Patent No. 6211164
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Ian, Yuo
; APPLICANT: Glenda, Vincent L.
; APPLICANT: Rockow-Magnone, Shayna K.
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES OF THE HUMAN
; FILE REFERENCE: 6675-US-01
; CURRENT APPLICATION NUMBER: US/09/522,800
; CURRENT FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-522-800-16

Query Match      78.8%; Score 26; DB 4; Length 476;
Best Local Similarity 71.4%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      1 TRIFSKL 7
      || |||
Db      378 TRFTKL 384

RESULT 15
US-08-924-183-1
; Sequence 1, Application US/08924183A
; Patent No. 6218109
; GENERAL INFORMATION:
; APPLICANT: Elledge, Stephen J.
; APPLICANT: Sanchez, Yolanda
; TITLE OF INVENTION: MAMMALIAN CHECKPOINT GENES AND PROTEINS
; FILE REFERENCE: 120541-1003
; CURRENT APPLICATION NUMBER: US/08/924,183A
; CURRENT FILING DATE: 1997-09-05
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-924-183-1

Query Match      78.8%; Score 26; DB 4; Length 476;
Best Local Similarity 71.4%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 TRIFSKL 7
      || |||
Db      378 TRFTKL 384
```

Search completed: July 15, 2002, 12:59:17  
Job time: 388 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:00:59 ; Search time 95.45 Seconds  
(without alignments)  
7.047 Million cell updates/sec

Title: us-09-712-819a-3

Perfect score: 35

Sequence: 1 FYQLALT 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pirl:\*  
2: pirl:\*  
3: pirl:\*  
4: pirl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	31	88.6	2	S50524
2	30	85.7	2	S45868
3	30	85.7	2	T16710
4	30	85.7	2	E95009
5	30	85.7	2	C97881
6	29	82.9	2	H64216
7	29	82.9	2	B82230
8	29	82.9	2	T43016
9	29	82.9	2	E71465
10	29	82.9	2	G81727
11	29	82.9	2	T50317
12	29	82.9	2	S54525
13	29	82.9	2	G75578
14	28	80.0	1	F69857
15	28	80.0	1	S44624
16	28	80.0	1	B83803
17	28	80.0	2	US0594
18	28	80.0	2	AG1873
19	28	80.0	2	T25506
20	28	80.0	2	H95902
21	28	80.0	2	S76598
22	28	80.0	2	H71482
23	28	80.0	2	G81745
24	28	80.0	2	T16084
25	28	80.0	1	MM87H2
26	28	80.0	2	S48313
27	28	80.0	2	S51081
28	28	80.0	2	T38362
29	28	80.0	2	S62408

30	28	80.0	783	2	AH2226	hypothetical prote
31	28	80.0	905	2	T23510	hypothetical prote
32	28	80.0	2130	2	AB0821	probable exported
33	27	77.1	89	2	I52695	cyclin-dependent k
34	27	77.1	105	2	D69619	multidrug resistan
35	27	77.1	153	2	S76676	hypothetical prote
36	27	77.1	221	2	S95922	hypothetical prote
37	27	77.1	240	1	Y1ZM91	19k zein precursor
38	27	77.1	240	1	Y1ZM92	19k zein precursor
39	27	77.1	240	1	Y1ZM93	19k zein precursor
40	27	77.1	240	2	T04364	19k zein precursor
41	27	77.1	240	2	A22831	hypothetical prote
42	27	77.1	249	2	E90048	hypothetical prote
43	27	77.1	250	2	H83784	glycerophospholies
44	27	77.1	267	2	T14824	alpha kafirin - so
45	27	77.1	269	2	T04095	ribonuclease S hom

#### ALIGNMENTS

RESULT 1  
S50524  
hypothetical protein YEL065w - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C>Date: 28-May-1993 #sequence\_revision 24-Feb-1995 #text\_change 23-Mar-2001  
C:Accession: S50524  
R:Dieckhoff, F.S.  
Submitted to the EMBL Data Library, December 1994  
A:Description: The sequence of S. cerevisiae cosmid 9669, 8334, 8199, and lambda clo  
A:Reference number: S50434  
A:Accession: S50524  
A:Molecule type: DNA  
A:Residues: 1628 <DIR>  
A:Cross-references: EMBL:U18795; NID:9603241; PIDD:AA065022.1; PID:9603253; GSPDB:GNO  
C:Gene: MIPS:YEL065w  
C:Gene: MIPS:YEL065w  
A:Map position: 5L  
C:Superfamily: conserved probable membrane protein YEL073c

Query Match 88.6% Score 31; DB 2; Length 628;  
Best local similarity 85.7% Pred. No. 26;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FYQLALT 7

DB 168 FYQLALT 174

RESULT 2  
S45868  
probable membrane protein YBR013c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YBR0209

C:Species: Saccharomyces cerevisiae

C:Date: 26-Aug-1994 #sequence\_revision 09-Sep-1994 #text\_change 04-Mar-2000

C:Accession: S45868

R:Entian, K.D.; Koetter, P.; Rose, M.; Li, Z.; Thermann, R.; Brendel, M.; Baur, A.; B

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45862

A:Accession: S45868

A:Molecule type: DNA

A:Residues: 1-129 <ENT>

A:Cross-references: EMBL:Z35862; NID:9577857; PID:9577860; GSPDB:GN00002; MIPS:YBR013

A:Experimental source: strain S288C

C:Genetics:

A:Gene: MIPS:YBR013c

A:Map position: 2R

C:Superfamily: Saccharomyces cerevisiae probable membrane protein YBR013c

C:Keywords: transmembrane protein

F:1-18/Domain: transmembrane #status predicted <TM>

Query Match 85.7%; Score 30; DB 2; Length 129;  
Best Local Similarity 85.7%; Pred. No. 8.2;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FYOLALT 7  
Db 120 FYOLALT 126

## RESULT 3

T16710

Hypothetical protein R10H1.2 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T16710

R:Bentley, D.

Submitted to the EMBL Data Library, July 1995

A:Description: The sequence of *C. elegans* cosmid R10H1.

A:Reference number: 218563

A:Accession: T16710

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-317 &lt;BEN&gt;

A:Cross-references: EMBL:023174; NID:g726398; PID:g726400; PIDN:MAC46710.1; CESP:R10H1.2

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:R10H1.2

A:Introns: 12/3; 65/3; 126/2; 182/3; 251/1; 279/3

## Query Match

Best Local Similarity 85.7%; Score 30; DB 2; Length 317;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FYOLALT 7  
Db 124 FYOLALT 130

## RESULT 4

E95009

Cell wall surface anchor family protein [Imported] - *Streptococcus pneumoniae* (strain T)C:Species: *Streptococcus pneumoniae*

C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: E95009

R:Uetzel, H.; Nelson, K.E.; Paulsen, I.T.; Elsen, J.A.; Read, T.D.; Peterson, S.; Held

on, J.D.; Umeyam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,

nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A:Title: Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: E95009

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-857 &lt;KUR&gt;

A:Cross-references: GB:AE005672; PIDN:AAK74270.1; PID:g14971548; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0082

## Query Match

Best Local Similarity 85.7%; Score 30; DB 2; Length 857;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FYOLALT 6  
Db 538 FYOLALT 543

## RESULT

C97881

conserved hypothetical protein spr0075 [Imported] - *Streptococcus pneumoniae* (strain  
C:Species: *Streptococcus pneumoniae*  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
C:Accession: C97881  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.  
e, R.; Leblanc, D.J.; Lee, L.N.; Letkovitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.  
A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: C97881  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1161 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAK98079.1; PID:g15457609; GSPDB:GN00174  
C:Genetics:

A:Gene: spr0075

## Query Match

Best Local Similarity 85.7%; Score 30; DB 2; Length 1161;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FYOLALT 6  
Db 842 FYOLALT 847

## RESULT 6

H64216

ribosomal protein l4 - *Mycoplasma genitalium*C:Species: *Mycoplasma genitalium*

C:Date: 17-Nov-1995 #sequence\_revision 17-Nov-1995 #text\_change 07-Dec-1999

C:Accession: H64216

R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.

M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,

C.A.; Venter, J.C.

Science 270, 397-403, 1995

A:Title: The minimal gene complement of *Mycoplasma genitalium*.

A:Reference number: A64200; MUID:96026346

A:Accession: H64216

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-211 &lt;TIGR&gt;

A:Cross-references: GB:U09695; GB:I43967; NID:g1045833; PID:g1045836; TIGR:MG152

A:Experimental source: strain G-37

C:Genetics:

A:Genetic code: SGC3

A:Superfamily: Escherichia coli ribosomal protein l4

## Query Match

Best Local Similarity 82.9%; Score 29; DB 2; Length 211;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 YOLALT 7  
Db 107 YOLALT 112

## RESULT 7

B82230

conserved hypothetical protein VC1194 [Imported] - *Vibrio cholerae* (strain N16961 serC:Species: *Vibrio cholerae*

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: B82230

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bess, S.; Qin, H.; Dragoi, I.; Sellers

L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A:Reference number: A82035; MUID:20406833



A:Accession: B82230  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-350 <HEI>  
A:Cross-references: GB:AE004199; GB:AE003852; NID:9655666; PIDN:AAF94353.1; GSPDB:GN001  
A:Experimental source: serogroup O1; strain N16961; biotype EI Tor  
C:Genetics:  
A:Gene: VC1194  
A:Map position: 1

Query Match 82.9%; Score 29; DB 2; Length 350;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 FYOLALT 7  
||| |||  
Db 50 FYOLALT 55

RESULT 8  
T43016  
DNA mismatch repair protein MTH1 homolog - fission yeast (Schizosaccharomyces pombe) (fr  
C:Species: Schizosaccharomyces pombe  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 02-Sep-2000  
C:Accession: T43016  
R:Yoshioka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.  
DNA Res. 4, 363-369, 1997  
A:Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.  
A:Reference number: 217323; MUID:98162722  
A:Accession: T43016  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-380 <YOS>  
A:Cross-references: EMBL:DB9212; NID:g1749631; PIDN:BA113873.1; PID:g1749632  
A:Experimental source: strain PR745  
C:Superfamily: mismatch repair protein MTH1

Query Match 82.9%; Score 29; DB 2; Length 380;  
Best Local Similarity 71.4%; Pred. No. 46;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FYOLALT 7  
||| |||  
Db 203 FYOLALT 209

RESULT 9  
E71465  
hypothetical protein CT837 - Chlamydia trachomatis (serotype D, strain UM3/Cx)  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 18-Aug-2000  
C:Accession: E71465  
R:Stephens, R.S.; Kaiman, S.; Lamme, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra  
A:Reference number: A71570; MUID:99000809  
A:Accession: E71465  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-658 <ARN>  
A:Cross-references: GB:AE001356; GB:AE001273; NID:g3329299; PIDN:AMC68434.1; PID:g332930  
A:Experimental source: serotype D, strain UM-3/Cx  
C:Genetics:  
A:Gene: CT837  
C:Superfamily: conserved hypothetical protein TC0225

Query Match 82.9%; Score 29; DB 2; Length 658;  
Best Local Similarity 85.7%; Pred. No. 83;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FYOLALT 7  
||| |||  
Db 217 FYOLALT 223

RESULT 10  
G81727  
conserved hypothetical protein TC0225 [imported] - Chlamydia muridarum (strain N199)  
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 18-Aug-2000  
C:Accession: G81727  
R:Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke, C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe, Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR391  
A:Reference number: A81500; MUID:20150255  
A:Accession: G81727  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-658 <TET>  
A:Cross-references: GB:AE002289; GB:AE002160; NID:g7190254; PIDN:AAF39097.1; PID:g7191  
A:Experimental source: strain N199 (MoPn)  
C:Genetics:  
A:Gene: TC0225  
C:Superfamily: conserved hypothetical protein TC0225

Query Match 82.9%; Score 29; DB 2; Length 658;  
Best Local Similarity 85.7%; Pred. No. 83;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FYOLALT 7  
||| |||  
Db 217 FYOLALT 223

RESULT 11  
T50317  
probable DNA mismatch repair protein, MTH1 homolog [imported] - fission yeast (Schizo  
C:Species: Schizosaccharomyces pombe  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 02-Sep-2000  
C:Accession: T50317  
R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Cadieu, E.; LeLaure, V.; Gilbert submitted to the EMBL Data Library, January 2000  
A:Reference number: Z25061  
A:Accession: T50317  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-684 <MCD>  
A:Cross-references: EMBL:AL136536; PIDN:CA66448.1; GSPDB:GN00067; SPDB:SPC1703.04  
A:Experimental source: strain 972h(-); cosmid cl703  
C:Genetics:  
A:Gene: SPDB:SPC1703.04  
A:Map position: 2  
A:Introns: 24/3; 70/3; 128/2  
C:Superfamily: mismatch repair protein MTH1

Query Match 82.9%; Score 29; DB 2; Length 684;  
Best Local Similarity 71.4%; Pred. No. 86;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FYOLALT 7  
||| |||  
Db 507 FYOLALT 513

RESULT 12  
S54525  
mismatch repair protein MTH1 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YMO520.16; protein YMK167  
C:Species: Saccharomyces cerevisiae  
C:Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 29-Sep-1999

C:Accession: S54525; S48890; S54612  
R: Hunt, S.; Bowman, S.  
submitted to the EMBL Data Library, May 1995  
A:Reference number: S54510  
A:Accession: S54525  
A:Molecule type: DNA  
A:Residues: 1-769 <HUN>  
A:Cross-references: GB:249705; EMBL:249700; NID:9825556; PIDN:CAA89803.1; PID:9825572; E  
A:Experimental source: strain AB972  
R:Prolla, T.A.; Christie, D.M.; Liskey, R.M.  
Mol. Cell. Biol. 14, 407-415, 1994  
A:Title: Dual requirement in yeast DNA mismatch repair for MLH1 and PMS1, two homologs  
A:Reference number: S48890; MUID:94088538  
A:Accession: S48850  
A:Molecule type: DNA  
A:Residues: 1-257, 'L', 259-287, 'F', 289-707, 'L', 709-769 <PRO>  
A:Cross-references: EMBL:007187; NID:9460626; PIDN:AAI16835.1; PID:9460627  
C:Genetics:  
A:Gene: SGD:MLH1  
A:Cross-references: SGD:S0004777; MIPS:YMR167w  
A:Map position: 13R  
C:Function:  
A:Description: DNA repair  
C:Superfamily: mismatch repair protein MLH1  
C:Keywords: DNA repair

Query Match 82.9%; Score 29; DB 2; Length 769;  
Best Local Similarity 71.4%; Pred. No. 98;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FYOALAT 7  
|||: ||  
Db 570 FYOIGLT 576

RESULT 13  
G75578  
aculeacin A acylase - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
A:Accession: G75578  
R:White, O.; Elsen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.; Ma  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: G75578  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-785 <WHI>  
A:Cross-references: GB:AE001863; GB:AE001825; NID:96460670; PIDN:AAF12385.1; PID:9646066  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DRA0255  
A:Map position: 2

Query Match 82.9%; Score 29; DB 2; Length 785;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FYOALAT 7  
|||||: ||  
Db 257 FYOALAT 263

RESULT 14  
F69857  
conserved hypothetical protein ykna - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000

C:Accession: F69857  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.; Bron, S.; Brouillet, S.; Busch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emmerston, P.T.; Entian, K.D.; Errington, J.; Fabbre, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M  
Ketter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
A:Authors: Laber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
Nleger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl  
A:Authors: Schlecht, S.; Schroeder, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Se  
akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchida  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
A:Reference number: A69580; MUID:98044033  
A:Accession: F69857  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-156 <RUN>  
A:Cross-references: GB:299110; GB:AL009126; NID:92633472; PIDN:CAB13174.1; PID:926336  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ykna  
C:Superfamily: hypothetical protein yaaJ

Query Match 80.0%; Score 28; DB 1; Length 156;  
Best Local Similarity 71.4%; Pred. No. 31;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FYOALAT 7  
|||: ||  
Db 132 FYOYTLT 138

RESULT 15  
S44624  
C50C3.5 protein - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 30-Jun-2001  
A:Accession: S44624  
R:Du, Z.  
submitted to the EMBL Data Library, May 1993  
A:Description: Sequence of the C. elegans cosmid C50C3.  
A:Reference number: S44626  
A:Accession: S44624  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-178 <FAV>  
A:Cross-references: EMBL:L14433; NID:9289649; PIDN:AAA27976.1; PID:9289657  
C:Genetics:  
A:Insertions: 51/3; 117/3; 144/2; 162/1  
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology  
C:Keywords: EF hand  
F:47-79/Domain: calmodulin repeat homology <EF1>  
F:83-115/Domain: calmodulin repeat homology <EF2>

Query Match 80.0%; Score 28; DB 2; Length 178;  
Best Local Similarity 71.4%; Pred. No. 35;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FYOALAT 7  
|||: |||  
Db 160 YRRLAT 166

Search completed: July 15, 2002, 13:01:01  
Job time: 467 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:23:31 ; Search time 44.9 Seconds  
(without alignments)  
6.036 Million cell updates/sec

Title: US-09-712-819A-3

Perfect score: 35

Sequence: 1 FVQLALF 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	31	88.6	YEHS_YEAST	P39980 saccharomyc
2	30	85.7	YBM3_YEAST	P38215 saccharomyc
3	30	85.7	YRP2_CAEEL	Q09344 caenorhabdl
4	29	82.9	NIFW_SYNP8	O07360 myocloccc
5	29	82.9	RL4_MYCGE	P47398 myocloccc
6	29	82.9	MLH1_YEAST	P38920 saccharomyc
7	28	80.0	YLJ5_CAEEL	P34368 caenorhabdl
8	28	80.0	PLG3_CAEEL	P29988 anabrena sp
9	28	80.0	Y074_SYNP3	Q55790 synechocyst
10	28	80.0	HXT2_YEAST	P23585 saccharomyc
11	28	80.0	HXT0_YEAST	P43581 saccharomyc
12	28	80.0	KMT2_KLULA	P53381 kluytomyc
13	28	80.0	YAT6_SCHPO	Q08761 schizosacch
14	27	77.1	EBR8_BACSU	O31792 bacillus su
15	27	77.1	YC24_GALISU	P35912 galdieria s
16	27	77.1	ZEAB_MAIZE	P04702 zea mays (m
17	27	77.1	ZEAT_MAIZE	P04703 zea mays (m
18	27	77.1	ZEAB_MAIZE	P06676 zea mays (m
19	27	77.1	ZEAB_MAIZE	P06677 zea mays (m
20	27	77.1	YU86_YEAST	P47153 saccharomyc
21	27	77.1	PK1_ASFB7	P42493 african swi
22	27	77.1	PK1_ASFB2	P34208 african swi
23	27	77.1	CDK4_HUMAN	P11802 homo sapien
24	27	77.1	CDK4_MOUSE	P30285 mus musculu
25	27	77.1	CDK4_PIG	P79432 sus scrofa
26	27	77.1	CDK4_RAT	P35426 rattus norv
27	27	77.1	PHIL_HAEIN	O51901 haemophilus
28	27	77.1	E2B_XYUFA	O99695 xyloella fas
29	27	77.1	EXON_HSVSA	Q01013 herpesvirus
30	27	77.1	ENGA_NEIGH	O87407 neisseria g
31	27	77.1	YC24_GUITH	O87407 guiliardia
32	27	77.1	YC24_CYAPA	P46260 cyanophora
33	27	77.1	YC24_ODOSI	P49530 odontella s

34	27	77.1	487	1	YC24_PORPU	P51240 porphyra pu
35	27	77.1	495	1	SUPF_ECOLI	P77522 escherichia
36	27	77.1	618	1	SUS5_RAT	O63008 rattus norv
37	27	77.1	643	1	SUS5_HUMAN	O92911 homo sapien
38	27	77.1	762	1	ACOCX_BOVIN	O29437 bos taurus
39	27	77.1	765	1	PELD_ECOLI	P32674 escherichia
40	27	77.1	949	1	SRB4_YEAST	P50104 saccharomyc
41	27	77.1	1376	1	CRB8_HUMAN	P82279 homo sapien
42	27	77.1	1638	1	DPO3_LACTA	O9cdt7 lactococcus
43	27	77.1	2039	1	CCH1_YEAST	P50077 saccharomyc
44	26	74.3	72	1	YVAF_ECOLI	P75820 escherichia
45	26	74.3	91	1	YVAF_VACCV	P20545 vaccinia vi

## ALIGNMENTS

RESULT	1	STANDARD	PRT	628 AA.
YEHS_YEAST				
ID	YEHS_YEAST			
AC	P39980;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	01-FEB-1995 (Rel. 31, Last annotation update)			
DE	Hypothetical 70.6 kDa protein in HXT8-CAN1 intergenic region.			
GN	YEL065W.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID:4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-S288C / AB972;			
RA	Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,			
RA	Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,			
RA	Chung E., Duncan M., Guzman E., Hartzell G., Hunkeler-Smith S.,			
RA	Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,			
RA	Mosedale D., Nakahara K., Namath A., Norgren R., Oelher P., Oh C.,			
RA	Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,			
RA	Taylor P., Wei Y., Yellon M., Botstein D., Davis R.W.;			
RL	Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).			
CC	-----			
CC	EMBL: U18795; AAB65022.1; -			
DR	SGD; S0000791; SITI.			
DR	Hypothetical protein; Transmembrane.			
KW	Transmem	68		
FT	Transmem	107	127	POTENTIAL.
FT	Transmem	132	152	POTENTIAL.
FT	Transmem	164	184	POTENTIAL.
FT	Transmem	194	214	POTENTIAL.
FT	Transmem	225	245	POTENTIAL.
FT	Transmem	285	305	POTENTIAL.
FT	Transmem	317	337	POTENTIAL.
FT	Transmem	354	374	POTENTIAL.
FT	Transmem	394	414	POTENTIAL.
FT	Transmem	420	440	POTENTIAL.
FT	Transmem	448	468	POTENTIAL.
FT	Transmem	488	508	POTENTIAL.
FT	Transmem	559	579	POTENTIAL.
FT	SEQUENCE	628 AA;	70561 MW;	39E711CFB61CF05 CRC64;

Query Match 88.6%; Score 31; DB 1; Length 628;  
Best Local Similarity 85.7%; Pred. No. 12;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYOLALT 7  
 168 FYOLALT 174

Db

RESULT 2  
 YBM3\_YEAST  
 ID YBM3\_YEAST STANDARD; PRT; 129 AA.  
 AC P38215;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 01-OCT-1994 (Rel. 35, Last annotation update)  
 DE Hypothetical 14.5 kDa protein in IPPI-Tp1 intergenic region  
 DE precursor.  
 GN YBR013C OR YBR0209.  
 OS Saccharomyces cerevisiae (Baker's yeast)  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RA [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288C;  
 RA Entian K.-D., Koetter P., Rose M., Li Z., Therman R., Brendel M.,  
 RA Baur A., Boles E., Miosga T., Schaff-Gerstenschlaeger I.,  
 RA Zimmermann F.K.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: Z35882; CAA84955.1; -  
 DR PIR: S45868; S45868.  
 DR SGD: S0000217; YBR013C.  
 KM Hypothetical protein; Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 1 129 HYPOTHETICAL PROTEIN YBR013C.  
 FT SEQUENCE 129 AA; 14481 MW; DBE90D47E09644E9 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 129;  
 Best Local Similarity 85.7%; Pred. No. 3.9;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYOLALT 7  
 120 FYOLALT 126

Db

RESULT 3  
 YRP2\_CAEEL  
 ID YRP2\_CAEEL STANDARD; PRT; 317 AA.  
 AC 009344;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Hypothetical 35.8 kDa protein R10H1.2 from chromosome II.  
 GN R10H1.2.  
 OS Caenorhabditis elegans.  
 OC Rhabditidae; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RA [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Bentley D.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: SOME, TO C.ELEGANS F27E5.5.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: U23174; AAC46710.1; -  
 DR WormPep: R10H1.2; CE02029.  
 DR InterPro: IPR000276; GPCR\_Rhodopsn.  
 DR InterPro: IPR004151; Sre.  
 DR Pfam: PF03125; Sre; 1.  
 DR PROSITE: PS50262; G\_PROTEIN\_RECIP\_F1\_2; UNKNOWN\_1.  
 KM Hypothetical protein; Transmembrane.  
 FT TRANSMEM 70 90 POTENTIAL.  
 FT TRANSMEM 105 125 POTENTIAL.  
 FT TRANSMEM 147 167 POTENTIAL.  
 FT TRANSMEM 191 211 POTENTIAL.  
 FT TRANSMEM 232 252 POTENTIAL.  
 FT TRANSMEM 292 312 POTENTIAL.  
 FT SEQUENCE 317 AA; 35770 MW; 88B98B32BF15D4E2 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 317;  
 Best Local Similarity 85.7%; Pred. No. 10;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYOLALT 7  
 124 FYHIALT 130

Db

RESULT 4  
 NIFW\_SYN8  
 ID NIFW\_SYN8 STANDARD; PRT; 111 AA.  
 AC 007360;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Nitrogenase stabilizing/protective protein nifw.  
 GN Nifw.  
 OS Synechococcus sp. (strain PCC 8801 / RP-1) (Cyanothace PCC 8801).  
 OC Bacteria; Cyanobacteria; Chroococcales; Cyanothecae.  
 OX NCBI\_TaxID=41431;  
 RA [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:99231861; Pubmed-10217509;  
 RA Huang T.-C., Lin R.-F., Chu M.-K., Chen H.-M.;  
 RT "Organization and expression of nitrogen-fixation genes in the aerobic  
 RT nitrogen-fixing unicellular cyanobacterium Synechococcus sp. strain  
 RP-1.";  
 RE Microbiology 145:743-753(1999).  
 CC -1- FUNCTION: May protect the nitrogenase Fe-Mo protein from oxidative  
 CC damage (By similarity).  
 CC -1- SUBUNIT: Homotrimer; associates with nifD (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE NIFW FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: AF003700; AAC35197.1; -  
 DR Nitrogen fixation.  
 KW NITROGEN FIXATION.  
 SQ SEQUENCE 111 AA; 12931 MW; C85B39697BFC310E CRC64;



```

RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almsrough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craton M., Dear S., Du Z., Durin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Lalster N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showken R.,
RA Sime M., Smaildon N., Smith A., Smith K., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstein L., Wilkinson-Sprat J.,
RA Woldman P.,
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: STRONG, TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L14433; AAA27976.1; -
DR PIR; S44624; S44624.
DR HSSP; P02593; ICMG.
DR Wormpep; C50C05.5; CE00121.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; efhand; 2.
DR SMART; SM00054; EFh; 2.
DR PROSITE; PS00018; EF_HAND; 2.
KW Hypothetical protein; Calcium-binding; Repeat.
FT CA_BIND 60 71 EF_HAND 1 (POTENTIAL).
FT CA_BIND 96 107 EF_HAND 2 (POTENTIAL).
SQ SEQUENCE 178 AA: 20022 MW; EBFACAC74D8C57BD CRC64;
OY 1 FYQALAT 7
Db 160 YVRLAT 166

Query Match 80.0%; Score 28; DB 1; Length 178;
Best Local Similarity 71.4%; Pred. No. 16;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYQALAT 7
Db 160 YVRLAT 166

RESULT 8
ID PYG3_ANASP STANDARD; PRT; 236 AA.
AC P29988;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Phycobilisome rod-core linker polypeptide pcg3 (L-RC 27.2).
GN PCG3 OR ALR0536.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_Taxid=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92077441; PubMed=1743523;
RA Bryant D.A., Stilwell V.L., Glauser M., Frank G., Sidler W.,
RA Zuber H.;
RT "A small multigene family encodes the rod-core linker polypeptides of
RT Anabaena sp. PCC7120 phycobilisomes.";
RL Gene 107:91-99(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneo T., Nakamura T., Wolk C.P., Kuritz T., Sasamoto S.,

```

```

RA Watanabe A., Irituchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
CC -1- FUNCTION: ROD-CORE LINKER PROTEIN REQUIRED FOR ATTACHMENT OF
CC PHYCOCYANIN TO ALLOPHYCOCYANIN IN CORES OF PHYCOBILISOMES.
CC -1- FUNCTION: LINKER POLYPEPTIDES DETERMINE THE STATE OF AGGREGATION
CC AND THE LOCATION OF THE DISC-SHAPE PHYCOBILIPROTEIN UNITS WITHIN
CC THE PHYCOBILISOME AND MODULATE THEIR SPECTROSCOPIC PROPERTIES IN
CC ORDER TO MEDIATE A DIRECTED AND OPTIMAL ENERGY TRANSFER.
CC -1- SUBUNIT: THE PHYCOBILISOME IS A HEMIDISCAL STRUCTURE THAT IS
CC COMPOSED OF TWO DISTINCT SUBSTRUCTURES: A CORE COMPLEX (THAT
CC CONTAINS PHYCOBILIPROTEINS) AND A NUMBER OF RODS RADIATING FROM
CC THE CORE.
CC -----
CC -1- SIMILARITY: TO OTHER PHYCOBILISOME LINKER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M80435; AAA22038.1; -
DR PIR; AP003582; BAB72494.1; -
DR EMBL; AP003582; BAB72494.1; -
DR PIR; JS0594; JS0594.
DR InterPro; IPR001297; PBS_linker_poly.
DR Pfam; PF00427; PBS_linker_poly; 1.
KW Phycobilisome; Photosynthesis; Multigene family; Complete proteome.
FT INT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 236 AA: 27074 MW; 54293B7074F147B9 CRC64;
OY 1 FYQALAT 7
Db 93 FYQALAT 99

Query Match 80.0%; Score 28; DB 1; Length 236;
Best Local Similarity 71.4%; Pred. No. 22;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYQALAT 7
Db 93 FYQALAT 99

RESULT 9
ID Y074_SYNY3 STANDARD; PRT; 480 AA.
AC Q55790;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 52.8 kDa protein SLR0074.
GN SLR0074.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_Taxid=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiyama M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- SIMILARITY: BELONGS TO THE UPF0051 (YCF2a) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D64004; BAA10542.1; -  
 DR InterPro: IPR000825; UPS0051.  
 DR Pfam: PF01458; UPF0051; 1  
 KM Hypothetical protein, complete proteome.  
 SO SEQUENCE 480 AA; 52795 MW; 066845CB17A0F33A CRC64;

Query Match 80.0%; Score 28; DB 1; Length 480;  
 Best Local Similarity 71.4%; Pred. NO. 47;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYOLALT 7  
 Db 328 FYSTALT 334

RESULT 10  
 HXT2\_YEAST STANDARD; PRT; 541 AA.  
 AC P23585; (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE High-affinity glucose transporter HXT2.  
 GN HXT2 OR YMR011W OR YMR270.15.  
 GN Saccharomyces cerevisiae (Baker's yeast).  
 OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Devlin K., Church C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: HIGH-AFFINITY GLUCOSE TRANSPORTER. IS ONLY INDISPENSABLE  
 FOR GROWTH ON LOW GLUCOSE-CONTAINING MEDIA, BECAUSE S.CEREVISIAE  
 POSSESSES OTHER SUGAR TRANSPORTERS.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- INDUCTION: REPRESSED AT HIGH GLUCOSE CONCENTRATIONS.  
 CC -1- MISCELLANEOUS: GLUCOSE TRANSPORT IS THOUGHT TO BE MEDIATED BY TWO  
 KINETICALLY DISTINCT SYSTEMS, A GLUCOSE-REPRESSIBLE HIGH-AFFINITY  
 SYSTEM AND A CONSTITUTIVE LOW-AFFINITY SYSTEM.  
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M33270; AAA34701.1; -  
 DR EMBL: 248513; CAA8528.1; -  
 DR PIR: S12200; MMBYH2.  
 DR SGD: S0004613; HXT2.  
 DR InterPro: IPR003663; Sugar-transporter.  
 DR InterPro: IPR003662; sub\_transporter.  
 DR Pfam: PF00083; sugar\_tr; 1.  
 DR PRINTS: PR00171; SUGRTNSPORT.  
 DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 1.  
 DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; 1.

KM Repeat; Transmembrane; Sugar transport; Transport; Glycoprotein;  
 KW Phosphorylation.  
 FT DOMAIN 1 49  
 FT TRANSMEM 50 70  
 FT DOMAIN 71 107  
 FT TRANSMEM 108 128  
 FT DOMAIN 129 134  
 FT TRANSMEM 135 155  
 FT DOMAIN 156 162  
 FT TRANSMEM 163 183  
 FT DOMAIN 184 192  
 FT TRANSMEM 193 213  
 FT DOMAIN 214 227  
 FT TRANSMEM 228 248  
 FT DOMAIN 249 327  
 FT TRANSMEM 328 347  
 FT TRANSMEM 348 350  
 FT TRANSMEM 351 371  
 FT DOMAIN 372 379  
 FT TRANSMEM 380 400  
 FT DOMAIN 401 418  
 FT TRANSMEM 419 439  
 FT DOMAIN 440 456  
 FT TRANSMEM 457 477  
 FT DOMAIN 478 484  
 FT TRANSMEM 485 505  
 FT DOMAIN 506 541  
 FT MOD\_RES 266 266  
 FT MOD\_RES 539 539  
 FT CARBOHYD 82 82  
 SO SEQUENCE 541 AA; 59840 MW; 6AEFC0A87391CA7 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 541;  
 Best Local Similarity 71.4%; Pred. NO. 54;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYOLALT 7  
 Db 198 FYOLALT 204

RESULT 11  
 HXT0\_YEAST STANDARD; PRT; 546 AA.  
 AC P43581;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hexose transporter HXT10.  
 GN HXT10 OR YMR011W.  
 GN Saccharomyces cerevisiae (Baker's yeast).  
 OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RL MEDLINE=95400292; PubMed=7670463;  
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,  
 RA Sasanna S.-I., Sasanna M., Tsuchiya Y., Soeda E., Yokoyama K.,  
 RA Yamazaki M., Tashiro H., Eki T.;  
 RT "Analysis of the nucleotide sequence of chromosome VI from  
 RT Saccharomyces cerevisiae.";  
 RT Nat. Genet. 10:261-268(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Barrell B.G., Church C., Rajandream M.A.;  
 RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: PROBABLE GLUCOSE TRANSPORTER.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D50617; BAA09227.1; -
DR EMBL: D31600; BAA06493.1; -
DR EMBL: Z46255; CAA86344.1; -
DR SCD: S0001883; HXT10.
DR InterPro: IPR003663; Sugar_transporter.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr.1.
DR PRINTS: PR00171; SUGARTRANSPORT.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1: 1.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2: 1.
KW Repeat: Transmembrane; Sugar transport; Glycoprotein.
FT DOMAIN 1 44
FT TRANSMEM 45 65
FT DOMAIN 66 100
FT TRANSMEM 101 121
FT DOMAIN 122 127
FT TRANSMEM 128 148
FT DOMAIN 149 158
FT TRANSMEM 159 179
FT DOMAIN 180 185
FT TRANSMEM 186 206
FT DOMAIN 207 220
FT TRANSMEM 221 241
FT DOMAIN 242 324
FT TRANSMEM 325 341
FT DOMAIN 342 347
FT TRANSMEM 348 365
FT TRANSMEM 366 372
FT TRANSMEM 373 393
FT DOMAIN 394 415
FT TRANSMEM 416 436
FT DOMAIN 437 453
FT TRANSMEM 454 474
FT DOMAIN 475 475
FT TRANSMEM 476 496
FT DOMAIN 497 546
FT CARBOHYD 75 75
SQ SEQUENCE 546 AA; 60662 MW; B2D8DCDBBEC396F CRC64;

Query Match 80.0%; Score 28; DB 1; Length 546;
Best Local Similarity 71.4%; Pred. No. 54;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FVQALAT 7
Db 191 FVQALMT 197

RESULT 12
KHT2_KLULA STANDARD; PRT; 566 AA.
AC P53387;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hexose transporter 2.
GN KHT2.
OS Kluyveromyces fragilis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_Taxid=28985;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-JA6;
RA MEDLINE-98028406; PubMed-9363776;
RA Weirich J., Goffin P., Kuger P., Ferrero I., Breunig K.D.;
RT "Influence of mutations in hexose-transporter genes on glucose
RT repression in Kluyveromyces fragilis."
RL Eur. J. Biochem. 249:248-257(1997).
CC -1- FUNCTION: PROBABLE GLUCOSE TRANSPORTER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z47080; CAA87389.1; -
DR InterPro: IPR003663; Sugar_transporter.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr.1.
DR PRINTS: PR00171; SUGARTRANSPORT.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1: 1.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2: 1.
KW Repeat: Transmembrane; Sugar transport; Glycoprotein.
FT DOMAIN 22 61
FT TRANSMEM 62 82
FT DOMAIN 83 112
FT TRANSMEM 113 133
FT DOMAIN 134 140
FT TRANSMEM 141 161
FT DOMAIN 162 166
FT TRANSMEM 167 187
FT DOMAIN 188 198
FT TRANSMEM 199 219
FT DOMAIN 220 233
FT TRANSMEM 234 254
FT DOMAIN 255 333
FT TRANSMEM 334 353
FT DOMAIN 354 357
FT TRANSMEM 358 378
FT DOMAIN 379 385
FT TRANSMEM 386 406
FT DOMAIN 407 428
FT TRANSMEM 429 449
FT DOMAIN 450 465
FT TRANSMEM 466 486
FT DOMAIN 487 492
FT TRANSMEM 493 513
FT DOMAIN 514 566
FT CARBOHYD 88 88
SQ SEQUENCE 566 AA; 62727 MW; 9784173EC375444 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 566;
Best Local Similarity 71.4%; Pred. No. 56;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FVQALAT 7
Db 204 FVQALMT 210

RESULT 13
YA76_SCHPO STANDARD; PRT; 668 AA.
AC O09761;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 75.9 kDa protein C24H6.06 in chromosome I.
GN SPAC24H6.06;
```



OS Schizosaccharomyces pombe (Pission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 RX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.,  
 RA Wood V.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z54142; CAA90850.2; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 668 AA; 75851 MW; 3EBBA1B3CC30B61 CRC64;

Query Match 80.08; Score 28; DB 1; Length 668;  
 Best Local Similarity 71.4%; Pred. No. 67;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 PYQALAT 7  
 |||||  
 Db 220 FYQALMT 226

RESULT 14  
 ID EBRB\_BACSU STANDARD; PRT; 105 AA.  
 AC 031792;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Multidrug resistance protein EBRB.  
 GN EBRB.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE SMALL MULTIDRUG RESISTANCE (SMR)  
 CC PROTEIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z99113; CAB13614.1; -  
 DR Subtilisin; B012585; ebrA.  
 DR InterPro: IPR000390; DUF7.  
 DR Pfam: PF00893; DUF7; 1.  
 KW Transmembrane; Transport; Complete proteome.  
 FT TRANSMEM 2 22  
 FT TRANSMEM 35 55 POTENTIAL.  
 FT TRANSMEM 57 77 POTENTIAL.  
 FT TRANSMEM 84 104 POTENTIAL.

SO SEQUENCE 105 AA; 11349 MW; 9EA0FB58A0EBDA6 CRC64;  
 Query Match 77.1%; Score 27; DB 1; Length 105;  
 Best Local Similarity 71.4%; Pred. No. 16;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 PYQALAT 7  
 |||||  
 Db 45 FYALSLT 51

RESULT 15  
 ID YC24\_GALSU STANDARD; PRT; 221 AA.  
 AC P35912;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein in alpha 3 region (ORF X) (Fragment).  
 GN YCF24.  
 OS Galdieria sulphuraria.  
 OC Chloroplast.  
 OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;  
 OC Galdieria.  
 RX NCBI\_TaxID=130081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=14-1-1 / ISOLATE 107.79/GDETTINGEN;  
 RX MEDLINE=94033298; PubMed=8219057;  
 RA Kostreva M., Zetsche K.;  
 FT "Organization of plastid-encoded ATPase genes and flanking regions  
 FT including homologues of infB and tsf in the thermophilic red alga  
 FT Galdieria sulphuraria".  
 RL Plant Mol. Biol. 23:67-76(1993).  
 CC -1- SIMILARITY: BELONGS TO THE UPP0051 (YCF24) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X67814; CAA48016.1; -  
 DR Mendel; 529; CYACA; YCF24.1.  
 DR InterPro: IPR000825; UPP0051.  
 DR Pfam: PF01458; UPP0051; 1.  
 KW Chloroplast; Hypothetical protein.  
 FT NON\_PEP 1  
 SQ SEQUENCE 221 AA; 24390 MW; 5CD796CD91BD6FCD CRC64;

Query Match 77.1%; Score 27; DB 1; Length 221;  
 Best Local Similarity 71.4%; Pred. No. 36;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 PYQALAT 7  
 |||||  
 Db 69 FYVALT 75

Search completed: July 15, 2002, 13:23:32  
 Job time: 148 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:22:40 ; Search time 172.49 Seconds  
(without alignments)  
7.020 Million cell updates/sec

Title: US-09-712-819A-3  
Perfect score: 35  
Sequence: 1 FYQALNF 7

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_PROTOZOA:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_VIRUS:\*  
16: SP\_BACTERIA:\*  
17: SP\_ARCHAEA:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	91.4	1346	16 Q98IR5	Q98IR5 rhizobium 1
2	30	85.7	370	2 O52357	O52357 Bacillus pu
3	30	85.7	744	8 Q9TUL3	Q9TUL3 villaria c
4	29	85.7	857	16 Q9TUL3	Q9TUL3 villaria c
5	29	82.9	111	2 O07360	O07360 streptococ
6	29	82.9	350	16 Q9KSR2	Q9KSR2 synchococ
7	29	82.9	380	3 P78862	P78862 schizosacch
8	29	82.9	658	16 O84844	O84844 chlamydia t
9	29	82.9	658	16 O9PL82	O9PL82 chlamydia t
10	29	82.9	684	16 Q9P7M6	Q9P7M6 schizosacch
11	29	82.9	755	5 Q9BKJ9	Q9BKJ9 dictyostell
12	29	82.9	785	16 Q9RYQ4	Q9RYQ4 dictyostell
13	28	80.0	156	16 O34598	O34598 bacillus su
14	28	80.0	179	16 Q9KBA2	Q9KBA2 bacillus su
15	28	80.0	183	5 Q9NLU3	Q9NLU3 leishmania
16	28	80.0	241	5 O01464	O01464 caenorhabdi

17	28	80.0	262	2 P95777	P95777 streptococ
18	28	80.0	317	16 Q92M57	Q92M57 rhizobium m
19	28	80.0	353	5 Q9TVR2	Q9TVR2 caenorhabdi
20	28	80.0	483	16 O84690	O84690 chlamydia t
21	28	80.0	483	16 Q9PLP3	Q9PLP3 chlamydia m
22	28	80.0	501	5 O19498	O19498 caenorhabdi
23	28	80.0	554	6 O02713	O02713 sus scrofa
24	28	80.0	688	8 Q9BAS8	Q9BAS8 buxus henry
25	28	80.0	695	8 Q9BAS7	Q9BAS7 buxus henry
26	28	80.0	701	8 Q9MVL7	Q9MVL7 abutilon hy
27	28	80.0	1020	5 Q9XU82	Q9XU82 caenorhabdi
28	27	77.1	100	2 P74951	P74951 vibrio para
29	27	77.1	105	2 Q9R910	Q9R910 bacillus su
30	27	77.1	119	11 Q9RIV8	Q9RIV8 rattus norv
31	27	77.1	132	1 Q977R1	Q977R1 uncultured
32	27	77.1	145	10 Q9AEN6	Q9AEN6 oryza sativ
33	27	77.1	153	16 O55862	O55862 synchocyst
34	27	77.1	166	10 Q9AX62	Q9AX62 oryza sativ
35	27	77.1	180	8 Q96815	Q96815 steleoneema
36	27	77.1	207	10 Q41882	Q41882 zea mays (m
37	27	77.1	227	5 Q9BL42	Q9BL42 caenorhabdi
38	27	77.1	238	10 Q41883	Q41883 zea mays (m
39	27	77.1	240	16 Q41884	Q41884 zea mays (m
40	27	77.1	249	16 Q99RH2	Q99RH2 staphylococ
41	27	77.1	250	16 Q9KXY1	Q9KXY1 bacillus ha
42	27	77.1	267	10 O82052	O82052 sorghum bic
43	27	77.1	269	10 P93641	P93641 zea mays (m
44	27	77.1	274	17 Q97V36	Q97V36 sulfolobus
45	27	77.1	276	12 O57125	O57125 cottonnail

## ALIGNMENTS

RESULT	ID	Q98IR5	PRELIMINARY;	PRG;	1346 AA.
AC	Q98IR5	01-OCT-2001 (TREMBLrel. 18, Created)			
DT	01-OCT-2001 (TREMBLrel. 18, Last sequence update)				
DT	01-OCT-2001 (TREMBLrel. 18, Last annotation update)				
DE	ML2286 PROTEIN.				
GN	ML2286.				
OS	Rhizobium loti (Mesorhizobium loti).				
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;				
CC	Phyllobacteriaceae; Mesorhizobium.				
OX	NCBI_Taxid=381.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=AF303097.				
RX	MEDLINE=21082950; PubMed-11214968.				
RA	Kaneke T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,				
RA	Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,				
RA	Kishida Y., Kiyokawa C., Kohara M., Matsunoto M., Matsuno A.,				
RA	Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,				
RA	Takeuchi C., Yamada M., Tabata S.;				
RT	*Complete genome structure of the nitrogen-fixing symbiotic bacterium				
RT	Mesorhizobium loti.				
RL	DNA Res. 7:331-338(2000).				
DR	EMBL; AP002999; BAB49451.1; .				
DR	InterPro; IPR003975; SHALchannel.				
DR	PRINTS; PR01497; SHALCHANNEL.				
KW	Complete proteome.				
SO	SEQUENCE 1346 AA; 148727 MW; 7D1288A3FC26E879 CRC64;				

QY	1 FYQALNF 7	91.4%;	Score 32;	DB 16;	Length 1346;
DB	696 FYQALNF 702	Best local Similarity	85.7%;	Pred: NO. 73;	
		Matches 6;	Conservative 1;	Mismatches 0;	Gaps 0;

```
RESULT 2
ID 052357 PRELIMINARY; PRT: 370 AA.
AC 052357;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUFATIVE RAP.
OS Bacillus pumilus (Bacillus mesentericus).
OC Plasmid pPL10.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
OX NCBI_TaxID=1408;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 12140;
RX MEDLINE=76260025; PubMed=821919;
RA Lovett P.S., Duval E.J., Kegans K.M.;
RT "Bacillus pumilus plasmid pPL10: properties and insertion into
RL J. Bacteriol. 127:817-828(1976).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 12140;
RA Lovett P.S., Rogers E.J.;
RT "Sequence of Bacillus pumilus plasmid pPL10."
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF036712; AAB91475.1; -.
DR InterPro: IPR001440; TPR.
DK Pfam: PF00515; TPR; 4.
KW Plasmid.
SQ SEQUENCE 370 AA; 44275 MW; 7302D7A53EBB3987 CRC64;

Query Match 85.7%; Score 30; DB 2; Length 370;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FYOLAL 6
DB 256 FYOLAL 261

RESULT 3
ID 09TL93 PRELIMINARY; PRT: 744 AA.
AC 09TL93;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT F.
GN NDHF.
OS Villarsia calthifolia.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Asterales; Menyanthaceae; Villarsia.
OX NCBI_TaxID=13756;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20340987; PubMed=10877943;
RA Olmstead R.G., Kilm K.J., Jansen R.K., Wagsstaff S.J.;
RT "The phylogeny of the asteridae sensu lato based on chloroplast ndhp
RT gene sequences."
RL Mol. Phylogenet. Evol. 16:96-112(2000).
CC -1- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE - NAD(+) + PLASTOQUINOL.
DR EMBL: AF130180; AAF08136.1; -.
DR InterPro: IPR001750; Oxidored_g1.
DR InterPro: IPR002128; Oxidored_g1_C.
DR InterPro: IPR001516; Oxidored_g1_N.
DK Pfam: PF00361; oxidored_g1; 1.
RN [1]
```

```
DR Pfam: PF001010; oxidored_g1_C; 1.
DR Pfam: PF00662; oxidored_g1_N; 1.
KW Chloroplast; NAD; Oxidoreductase; Plastoquinone.
SQ SEQUENCE 744 AA; 84557 MW; C98CD4FCCAD23724 CRC64;

Query Match 85.7%; Score 30; DB 8; Length 744;
Best Local Similarity 71.4%; Pred. No. 1.2e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 FYOLALT 7
DB 671 FYEALT 677

RESULT 4
ID 09TT70 PRELIMINARY; PRT: 857 AA.
AC 09TT70;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELL WALL SURFACE ANCHOR FAMILY PROTEIN.
GN SP0082.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwin M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anguillo S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR FMBL: AE007325; AAK74270.1; -.
DR TIGR: SP0082; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 857 AA; 90923 MW; 987FEBFBCDE86F CRC64;

Query Match 85.7%; Score 30; DB 16; Length 857;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FYOLAL 6
DB 538 FYOLAL 543

RESULT 5
ID 007360 PRELIMINARY; PRT: 111 AA.
AC 007360;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NTFW.
GN NTFW.
OS Streptococcus sp. (strain PCC 8801 / RP-1) (Cyanothecae PCC 8801).
OC Bacteria; Cyanobacteria; Chroococcales; Cyanothecae.
OX NCBI_TaxID=41431;
RN [1]
```

RP SEQUENCE FROM N.A.  
 RC STRAIN-RE-1;  
 RX MEDLINE-99231861; PubMed-10217509;  
 RA Huang T.C., Lin R.F., Chu M.K., Chen H.M.;  
 RT "organization and expression of nitrogen-fixation genes in the aerobic  
 RT nitrogen-fixing unicellular cyanobacterium *Synechococcus* sp. strain  
 RT R-1."  
 RL Microbiology 145:743-753(1999).  
 DR EMBL: AF003700; AAC3197.1; -.  
 SQ SEQUENCE 111 AA; 12931 MW; C85B39697EFC310E CRC64;

Query Match 82.9%; Score 29; DB 2; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 YOLALT 7  
 DB 67 YOLALT 72

RESULT 6

ID O9KSR2 PRELIMINARY; PRT; 350 AA.  
 AC O9KSR2;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL PROTEIN VC1194.  
 GN VC1194.  
 OS *Vibrio cholerae*.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.  
 OX NCBI\_TaxID-666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE-20406833; PubMed-10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.R., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.F., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Olin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*  
 RT *cholerae*."  
 RL Nature 406:477-483(2000).  
 DR EMBL: AE004199; AAP94353.1; -.  
 DR TIGR: VC1194.  
 DR InterPro: IPR001623; DnaJ\_N.  
 DR SMART: SM00271; DnaJ; 1.  
 DR PROSITE: PS50076; DnaJ\_2; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 350 AA; 40627 MW; 6B6F035EAC79BBDE CRC64;

Query Match 82.9%; Score 29; DB 16; Length 350;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 YOLALT 7  
 DB 50 YOLALT 55

RESULT 7

ID P78862 PRELIMINARY; PRT; 380 AA.  
 AC P78862;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE UNKNOWN PROTEIN (FRAGMENT).

OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID-4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PR745;  
 RX MEDLINE-98162722; PubMed-9501991;  
 RA Yoshioke S., Kato K., Nakai K., Okayama H., Nojima H.;  
 RT "Identification of open reading frames in *Schizosaccharomyces pombe*  
 RT *cdnas*."  
 RL DNA Res. 4:363-369(1997).  
 DR EMBL: D89212; BAA13873.1; -.  
 FT NON TER 1  
 SQ SEQUENCE 380 AA; 43661 MW; 9AE19FB26E0A6303 CRC64;

Query Match 82.9%; Score 29; DB 3; Length 380;  
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYOLALT 7  
 DB 203 FYOLALT 209

RESULT 8  
 ID O84844 PRELIMINARY; PRT; 658 AA.  
 AC O84844;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 76.5 KDA PROTEIN.  
 GN C7837.  
 OS *Chlamydia trachomatis*.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; *Chlamydia*.  
 OX NCBI\_TaxID-813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-D/DW-3/CX;  
 RX MEDLINE-99000809; PubMed-9784136;  
 RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
 RA Mitchell W.P., Olinger L., Yatsusov R.L., Zhao Q., Koonin E.V.,  
 RA Davis R.W.;  
 RT "Genome sequence of an obligate intracellular pathogen of humans:  
 RT *Chlamydia trachomatis*."  
 RL Science 282:754-759(1998).  
 DR EMBL: AE001356; AAC68434.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 658 AA; 76502 MW; E17FC45063CCE717 CRC64;

Query Match 82.9%; Score 29; DB 16; Length 658;  
 Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYOLALT 7  
 DB 217 FYOLALT 223

RESULT 9

ID O9PL82 PRELIMINARY; PRT; 658 AA.  
 AC O9PL82;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL PROTEIN TC0225.  
 GN TC0225.  
 OS *Chlamydia muridarum*.

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83560;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MDPN / NIGG:  
RX MEDLINE=20150255; Pubmed=10684935;  
RA Read F.D., Bunham R.C., Shen C., Gill S.R., Heldelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
RA Linher K., Weidman J., Khouli H., Craven B., Bowman C., Dodson R.,  
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.,  
RT "Genome sequences of Chlamydia trachomatis MDPN and Chlamydia  
RT Pneumoniae AR39."  
RL Nucleic Acids Res. 28:1397-1406(2000).  
DR EMBL: AE002289; AAF39097.1; -.  
DR TIGR: TC0225; -.  
DR InterPro: IPR002129; Pyridoxal\_dec.  
DR PROSITE: PS00392; DDC\_GAD\_HDC\_YDC; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 658 AA; 76487 MW; BDD28A01C6DBCB8 CRC64;

Query Match 82.9%; Score 29; DB 16; Length 658;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 FYQALAT 7  
DB 217 FYQALAT 223

RESULT 10  
Q9P7W6 PRELIMINARY; PRT; 684 AA.  
AC Q9P7W6;  
DT 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
DE PUTATIVE DNA MISMATCH REPAIR PROTEIN, MLHI HOMOLOG.  
GN SPEC1703.04.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972H-;  
RA McDougall R.C., Rajandream M.A., Barrell B.G., Cadieu E., Lelaure V.,  
RA Gallibert F.,  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL136536; CAB66448.1; -.  
DR HSSP: P23367; 1BKN.  
DR InterPro: IPR002099; DNA\_mis\_repair.  
DR InterPro: IPR003594; HATPase\_C.  
DR Pfam: PF01119; DNA\_mis\_repair; 1.  
DR Pfam: PF02518; HATPase\_C; 1.  
DR PROSITE: PS00058; DNA\_MISMATCH\_REPAIR\_1; UNKNOWN\_1.  
SQ SEQUENCE 684 AA; 77253 MW; 7A59633B38E2944B CRC64;

Query Match 82.9%; Score 29; DB 3; Length 684;  
Best Local Similarity 71.4%; Pred. No. 1.9e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FYQALAT 7  
DB 507 FYQALAT 513  
RESULT 11  
Q9BKJ9 PRELIMINARY; PRT; 755 AA.

AC Q9BKJ9;  
DT 01-JUN-2001 (TREMblrel. 17, Created)  
DT 01-JUN-2001 (TREMblrel. 17, last sequence update)  
DT 01-OCT-2001 (TREMblrel. 18, last annotation update)  
DE LYSOSOMAL INTEGRAL MEMBRANE PROTEIN II.  
GN LMPB.  
OS Dictyostellium discoideum (slime mold).  
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellum.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Janssen K.-P., Schleicher M.,  
RT "Characterization of CD36/LIMP-II-Homologues in Dictyostellum  
RT discoideum."  
RL EMBL: AF238324; AAK30040.1; -.  
DR InterPro: IPR002159; CD36.  
DR Pfam: PF01130; CD36; 1.  
SQ SEQUENCE 755 AA; 82571 MW; 93BB639461C97547 CRC64;

Query Match 82.9%; Score 29; DB 5; Length 755;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 YQALAT 7  
DB 297 YQALAT 302

RESULT 12  
Q9RYO4 PRELIMINARY; PRT; 785 AA.  
AC Q9RYO4;  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, last annotation update)  
DE ACULACTIN A ACTINASE.  
GN DRA0255.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R1;  
RX MEDLINE=20036896; Pubmed=10567266;  
RA White O., Eisen J.A., Heldelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J., Lam P., McDonald L., Utterback T., Zaleski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Miron K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.,  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans R1."  
RL Science 286:1571-1577(1999).  
DR EMBL: AE001863; AAF12385.1; -.  
DR TIGR: DRA0255; -.  
DR InterPro: IPR002692; Penicill\_amidase.  
DR Pfam: PF01804; Penicill\_amidase; 2.  
KW Complete proteome.  
SQ SEQUENCE 785 AA; 84042 MW; 031166EAB2852778 CRC64;

Query Match 82.9%; Score 29; DB 16; Length 785;  
Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 FYQALAT 7  
DB 257 FYQALAT 263

```

RESULT 13
ID 034598 PRELIMINARY; PRT: 156 AA.
AC 034598;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE YKOA.
GN YKOA OR YKMA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boutsier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Enlian K.D., Erlington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly J., Harwood C.R., Henaut A.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hult M.F., Itaya M., Jones L.,
RA Joris B., Karamela D., Kasahara Y., Kieffer-Blochard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapilus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Meliado R.P., Mizuno M., Moestl D., Nakai S., Nock M.,
RA Moore D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scallan E., Schleich S., Schroeter R., Scottone F.,
RA Sekiguchi J., Sekowska A., Serot S.J., Serior P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Takakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
RA Viari A., Wambler R., Wedler E., Wedler H., Welltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.R., Zumbstein E., Yoshikawa H., Danchin A.;
RT *The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis.*;
RN Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ002571; CAA05596.1; -.
DR EMBL: Z99110; CAH13174.1; -.
DR InterPro: IPR001215; dCMP_cyt_dcam.
DR Pfam: PF00383; dCMP_cyt_dcam.1.
DR PROSITE: PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 156 AA; 17156 MW; B6498345A98BC214 CRC64;

```

```

Query Match 80.0%; Score 28; DB 16; Length 156;
Best Local Similarity 71.4%; Pred. No. 78;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 FYOLALT 7
|||||

```

```

Db 132 FYOVLTF 138
RESULT 14
ID 09KBA2 PRELIMINARY; PRT: 179 AA.
AC 09KBA2;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE RNA POLYMERASE ECF-TYPE SIGMA FACTOR.
GN BH2026.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86655;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT *Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.*;
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001514; BAB05745.1; -.
KW Complete proteome.
SQ SEQUENCE 179 AA; 21257 MW; E00A9640AC4D854A CRC64;

```

```

Query Match 80.0%; Score 28; DB 16; Length 179;
Best Local Similarity 71.4%; Pred. No. 90;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 FYOLALT 7
|||
Db 29 YRRLALT 35

```

```

RESULT 15
ID 09NLU3 PRELIMINARY; PRT: 183 AA.
AC 09NLU3;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE POSSIBLE PROLINE OXIDASE, MITOCHONDRIAL (FRAGMENT).
GN LM26.349.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL160493; CAB97967.1; -.
FT NON-TER 183
SQ SEQUENCE 183 AA; 20924 MW; 9C5FB37BB8F1AB2A CRC64;

```

```

Query Match 80.0%; Score 28; DB 5; Length 183;
Best Local Similarity 83.3%; Pred. No. 92;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 FYOLALT 6
|||||
Db 57 FYOLALT 62

```

Search completed: July 15, 2002, 13:22:42

Mon Jul 15 13:58:24 2002

us-09-712-819a-3.open.rspt

Page 6

Job time: 1478 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 15, 2002, 12:57:55 ; Search time 228.39 Seconds  
(without alignments)  
3.404 Million cell updates/sec

Title: US-09-712-819a-3  
Perfect score: 35  
Sequence: 1 FVQLALT 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: A\_Geneseq\_032802.\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT.\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT.\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT.\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT.\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT.\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT.\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT.\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT.\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT.\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT.\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT.\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT.\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT.\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT.\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT.\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT.\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT.\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT.\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT.\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT.\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT.\*  
23: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	88.6	751	22	AAE09958 Methylinomas 16a n
2	30	85.7	262	21	AAE09958 Human colon cancer
3	30	85.7	313	22	AAE09958 Human olfactory re
4	30	85.7	624	21	AAE09958 Streptococcus pneu
5	29	82.9	9	14	AAE09958 Residues 132-140 o
6	29	82.9	9	18	AAE09958 Human p53 peptide
7	29	82.9	10	18	AAE09958 Antigenic human p5
8	29	82.9	14	20	AAE09958 Mutant P53 tumour
9	29	82.9	35	20	AAE09958 Mutant P53 tumour
10	29	82.9	64	22	AAE09958 Human colon cancer
11	29	82.9	491	20	AAE09958 Protein which is s

12	29	82.9	769	22	AAE09958 Yeast MHL1 protein
13	29	82.9	769	22	AAE09958 Amino acid sequenc
14	29	82.9	3391	12	AAE09958 Proteins encoded b
15	28	80.0	247	20	AAE09958 Chlamydia trachoma
16	28	80.0	247	22	AAE09958 Escherichia coli p
17	28	80.0	480	21	AAE09958 Synechocystis sp.
18	28	80.0	1020	22	AAE09958 C. elegans RNA int
19	27	77.1	16	21	AAE09958 Consensus site for
20	27	77.1	22	19	AAE09958 Human neurofilamen
21	27	77.1	124	22	AAE09958 Human colon cancer
22	27	77.1	240	11	AAE09958 Amino acid sequenc
23	27	77.1	240	21	AAE09958 Human vesicle asso
24	27	77.1	240	22	AAE09958 Human AFP protein
25	27	77.1	240	22	AAE09958 Human type I membr
26	27	77.1	267	21	AAE09958 Human ORF2598
27	27	77.1	278	20	AAE09958 Human protein kina
28	27	77.1	303	21	AAE09958 Lung cancer associ
29	27	77.1	303	16	AAE09958 CDK4 R24C mutant.
30	27	77.1	303	20	AAE09958 Mouse CDK-4 protei
31	27	77.1	303	21	AAE09958 Cyclin-dependent X
32	27	77.1	303	21	AAE09958 Human cyclin depen
33	27	77.1	354	22	AAE09958 Human organic cati
34	27	77.1	354	22	AAE09958 Human protein SBO
35	27	77.1	399	21	AAE09958 Arabidopsis thalia
36	27	77.1	431	21	AAE09958 Arabidopsis thalia
37	27	77.1	484	20	AAE09958 Chlamydia pneumoni
38	27	77.1	485	20	AAE09958 Human pp392.3 secr
39	27	77.1	485	22	AAE09958 Human secreted pro
40	27	77.1	508	21	AAE09958 Escherichia coli Y
41	27	77.1	555	22	AAE09958 Human protein SBO
42	27	77.1	557	21	AAE09958 Arabidopsis thalia
43	27	77.1	618	18	AAE09958 Rat sodium/Iodide
44	27	77.1	618	18	AAE09958 Human cyclin D1-hu
45	27	77.1	618	19	AAE09958 Human cyclin D1/cy

#### ALIGNMENTS

RESULT	ID	AAE09958	standard; Protein; 751 AA.
1	XX	AAE09958	
29-NOV-2001 (first entry)	XX	AAE09958	
Methylinomas 16a nitric oxide reductase NorZ.	XX		
Denitrification; enzyme; bacteria; fermentation; nitric oxide reductase;	XX		
NorZ.	XX		
Methylinomas sp. 16a.	OS		
WO200164898-A2.	XX		
07-SEP-2001.	PN		
22-FEB-2001; 2001MO-US05900.	XX		
29-FEB-2000; 2000US-0185621.	XX		
(DUPO ) DU PONT DE NEMOURS & CO E I.	XX		
Odum JM, Norton KC, Schenzle AJ, Ye RW;	PA		
WPI: 2001-557773/62.	PI		
N-PSDB: AAD17012.	XX		
Novel nucleic acid fragments from Methylinomas species for use in	XX		
bacterial denitrification process; encodes bacterial nitrite reductase,	XX		
bacterial nitric oxide reductase or bacterial nitrate reductase	XX		

PS Claim 17; Page 60-62; 64pp; English.

XX The invention relates to a nucleic acid fragment from *Methylobacter*

CC encoding nitrite reductase, nitric oxide reductase or nitrate reductase.

CC The nucleic acid fragment is isolated from a gene cluster containing 11

CC open reading frames (ORFs) encoding enzymes involved in denitrification.

CC The enzymes are useful in microbial denitrification reactions, for

CC identifying other denitrifying bacteria and for fermentation processes

CC in absence or presence of oxygen. Chimeric genes comprising the

CC polynucleotide of the invention can be used to alter the level of

CC denitrifying enzyme in the recombinant host. The polynucleotide is

CC useful for isolating genes encoding homologous proteins from other

CC microbial species, for expression of various pathway intermediates, for

CC modulation of pathway already existing in the host for the synthesis of

CC new products, and for conferring higher growth yields on the host or for

CC enabling alternative growth mode to be utilised. The present sequence is

CC nitric oxide reductase No2r from *Methylobacter* 16a.

XX

SQ Sequence 751 AA;

Query Match 88.6%; Score 31; DB 22; Length 751;

Best Local Similarity 85.7%; Pred. No. 94;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYQALR 7

Db 501 fyqalgt 507

RESULT 2

AAB53254

ID AAB53254 standard; Protein: 262 AA.

XX

XX AAB53254;

XX

XX 09-MAR-2001 (first entry)

XX

DE Human colon cancer antigen protein sequence SEQ ID NO:794.

XX

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX identification; cytostatic; cardioactive; neuroprotective; vulnerrary;

XX immunomodulatory; muscular; gynaecological; gastrointestinal;

XX nephrotoxic; antiinfective; antibacterial; gene therapy; wound;

XX neural disorder; immune system disorder; muscular disorder;

XX reproductive disorder; gastrointestinal disorder; renal disorder;

XX infectious disease; cardiovascular disorder.

XX

OS Homo sapiens.

XX

XX

PN MO200055351-A1.

XX

XX 21-SEP-2000.

XX

XX 08-MAR-2000; 2000MO-US05883.

XX

XX 12-MAR-1999; 99US-0124270.

XX

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

XX

PI Rosen CA, Ruben SM;

XX

XX WPI: 2000-587534/55.

XX

XX N-PSDB; AAC98011.

XX

XX Colon cancer associated gene sequences, referred to as colon cancer

XX PT antigens, useful for the treatment, prevention, and diagnosis of colon

XX disorders such as colon cancer -

XX

PS Claim 11; Page 1350-1351; 2104pp; English.

XX

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,

XX CC called human colon cancer antigens, given in AAB53234 to AAB54006. The

CC human colon cancer antigens can have cytostatic, cardioactive, muscular;

CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,

CC vulnerrary, nephrotoxic, antiinfective and antibacterial activities, and

CC can be used in gene therapy. The colon cancer antigen polynucleotides,

CC proteins and antibodies to the proteins are useful for the prevention,

CC treatment and diagnosis of colon disorders, such as colon cancer. The

CC polynucleotides may be used in diagnostics and research, such as for

CC chromosome identification, and as hybridisation probes. The proteins

CC may also be used to prevent diseases such as neural disorders, immune

CC system disorders, muscular disorders, reproductive disorders,

CC gastrointestinal disorders, wounds, renal disorders, infectious

CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and

CC AAB54007 represent sequences used in the exemplification of the present

XX invention.

XX

SQ Sequence 262 AA;

Query Match 85.7%; Score 30; DB 21; Length 262;

Best Local Similarity 100.0%; Pred. No. 51;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYQALR 6

Db 19 fyqal 24

RESULT 3

AAG71826

ID AAG71826 standard; Protein: 313 AA.

XX

XX AAG71826;

XX

XX 31-JUL-2001 (first entry)

XX

DE Human olfactory receptor polypeptide, SEQ ID NO: 1507.

XX

XX Human; olfactory receptor; OR; primary scent determination;

XX secondary scent determination; polypeptide library; odour receptor;

XX scent profile; scent fingerprint; scent representation.

XX

OS Homo sapiens.

XX

XX

PN MO200127158-A2.

XX

XX 19-APR-2001.

XX

XX 06-OCT-2000; 2000MO-US27582.

XX

XX

XX 08-OCT-1999; 99US-0158615.

XX

XX 24-FEB-2000; 2000US-0184809.

XX

XX (DIGI-) DIGISCENTS.

XX (YEDA ) YEDA RES & DEV CO LTD.

XX

PI Bellenson J, Smith D, Lancel D, Glusman G, Fuchs T, Yanai I;

XX

XX WPI: 2001-290713/30.

XX

XX

XX New polynucleotides which encode polypeptides involved in olfactory

XX PT sensation for identifying olfactory agonists and antagonists -

XX

XX

PS Claim 11; Page 959-960; 1857pp; English.

XX

XX The present sequence is an olfactory receptor which is encoded by

XX CC one of a number of novel polynucleotides. The polynucleotides can be

XX CC used in screening for olfactory agonists and antagonists. The methods

XX CC allow for the determination of primary scents and the identification

XX CC of the odour receptors used to detect these primary scents. The methods

XX CC also enable determination of secondary scents and the identification of

XX CC combinations of odour receptors that are involved in detecting such

XX CC secondary scents. This enables the construction of a scent representation

XX CC (also called a scent fingerprint or scent profile), which may be used to

CC re-create and edit scents. Libraries of olfactory receptors are useful  
 CC for determining the interaction pattern of a composition with the  
 CC receptors, and can be used for determining differences in the olfactory  
 CC faculties of different individuals.

XX Sequence 313 AA;

Query Match 85.7%; Score 30; DB 22; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYOLAL 6  
 Db 197 fyqlal 202

#### RESULT 4

AAV81588 standard; Protein; 624 AA.

XX AAV81588;

DT 24-MAY-2000 (first entry)

XX Streptococcus pneumoniae type 4 protein sequence #88.

XX Streptococcus pneumoniae; vaccine; screening; protein antigen;  
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;  
 KW pneumococcal disease.

XX Streptococcus pneumoniae.

XX WO200006737-A2.

XX 10-FEB-2000.

XX 27-JUL-1999; 99WO-GB02451.

XX 27-JUL-1998; 98GB-0016337.

PR 19-MAR-1999; 99US-0125164.

XX (MCCR-) MICROBIAL TECHNIQS LTD.

XX Glibert CPG, Hansbro PM;

DR WPI; 2000-195300/17.

PT New Streptococcal protein, useful as a vaccine, for diagnosis of  
 PT pneumococcal diseases and for screening agents capable of antagonizing  
 PT or inhibiting expression of the protein

PS Claim 1; Page 81; 108pp; English.

XX AAV81501 to AAV81679 represent specifically claimed protein sequences  
 CC isolated from Streptococcus pneumoniae. AAV815407 to AAV815590 represent  
 CC specifically claimed nucleotide sequences isolated from S. pneumoniae.  
 CC The sequences have antibacterial and antiinflammatory properties.  
 CC The protein sequences, and fragments of them, are useful as immunogens  
 CC and/or antigens. The nucleotide sequences can be used in vaccines and in  
 CC diagnostic assays. The proteins and nucleotides can be useful for the  
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also  
 CC useful for screening an agent capable of antagonizing, inhibiting or  
 CC interfering with the function or expression of the proteins in which the  
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection  
 CC and meningitis. AAV815591 to AAV815614 represent primers used in the  
 CC exemplification of the present invention.

XX Sequence 624 AA;

Query Match

Best Local Similarity 85.7%; Score 30; DB 21; Length 624;  
 100.0%; Pred. No. 1,3e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 FYOLAL 6  
 Db 538 fyqlal 543

#### RESULT 5

AAV4285 standard; peptide; 9 AA.

XX AAV4285;

DT 09-JUN-1994 (first entry)

XX Residues 132-140 of human mp53.

XX Human; p53; protein; Class I; allele; diagnosis; assay; detection;  
 KW p53 protein-specific; T-cell; antibody; cancer; MHC; HLA-A2.1.

XX Homo sapiens.

XX WO9324525-A.

XX 09-DEC-1993.

XX 18-MAY-1993; 93WO-MU00102.

XX 26-MAY-1992; 92EP-0201510.

XX (UYLE-) RIKSUNITIV LEIDEN.  
 PA (SEED-) SEED CAPITAL INVESTMENTS (SCI) BV.

XX Kast WM, Melief CJM;

XX WPI; 1993-405730/50.

XX Peptide derived from p53 protein - used to treat and diagnose  
 PT diseases involving over-expression of p53 e.g. human cancers

XX Claim 3; Page 38; 46pp; English.

XX The sequences given in AAV4281-94 are peptides derived from human mp53  
 CC protein which have the ability to bind to a human Class I molecule.  
 CC These peptides may be used in a diagnostic test or assay to detect  
 CC human p53 protein-specific T-cells or antibodies. They may also be  
 CC used in the treatment of diseases such as human cancers showing p53  
 CC protein overexpression. These peptides can bind human MHC Class I  
 CC allele HLA-A2.1.

XX Sequence 9 AA;

Query Match 82.9%; Score 29; DB 14; Length 9;  
 Best Local Similarity 85.7%; Pred. No. 6,4e+05;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYOLAL 7  
 Db 3 fyqlal 9

#### RESULT 6

AAV39629 standard; peptide; 9 AA.

XX AAV39629;

DT 11-JUN-1998 (first entry)

XX Human p53 peptide #6 (pos. 132-140).

XX T cell epitope; immune response; human leukocyte antigen; HLA Class I;

KW vaccine; immunogenic; major histocompatibility complex; MHC; B cell;  
 KW disease; anti-tumour; anti-viral.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO9741440-A1.  
 XX  
 PD 06-NOV-1997.  
 XX  
 PF 28-APR-1997; 97WO-NL00229.  
 XX  
 PR 23-DEC-1996; 96EP-0203670.  
 PR 26-APR-1996; 96EP-0201145.  
 XX  
 PA (UYLE-) RIKSUNIV LEIDEN.  
 PA (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.  
 XX  
 PI Kast WM, Mellef CJM, Offringa R, Toes REM, Van Der Burg SH;  
 XX  
 DR WPI: 1997-549891/50.  
 XX  
 PT Method of selecting T cell peptide epitope(s) - by measuring the  
 XX  
 PS stability of HLA class I-peptide complexes on intact B cells  
 XX  
 XX Example 3; Page 76; 109pp; English.  
 CC Peptides AAW39430-W29734 are used in a novel method for the selection of  
 CC immunogenic T-cell peptide epitopes present in polypeptide antigens. The  
 CC method involves the identification of peptide sequences capable of  
 CC binding to an HLA (human leukocyte antigen) class I molecule and  
 CC measuring the binding of this epitope peptide to the HLA class I  
 CC peptide. The stability of binding of the peptide and MHC (major  
 CC histocompatibility complex) class I molecule is measured on intact human  
 CC B cells carrying the MHC molecule at their cell surfaces. The method can  
 CC be used to select peptide epitopes for generating vaccines against a  
 CC disease associated with the polypeptide, e.g. cancers or AIDS. The  
 CC peptide epitopes are especially T-cell peptide epitopes with strong  
 CC anti-tumour and anti-viral immune responses. Peptide AAW39629 is derived  
 CC from human p53 and is capable of binding to the human MHC Class I allele  
 CC HLA-A2.1.  
 CC  
 XX  
 SQ Sequence 9 AA;  
 Query Match 82.9%; Score 29; DB 18; Length 9;  
 Best Local Similarity 85.7%; Pred. No. 6.4e+05;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 FYQLALT 7  
 | | | | |  
 Db 3 FYQLALT 9  
 RESULT 7  
 AAW22030  
 ID AAW22030 standard; peptide: 10 AA.  
 XX  
 AC AAW22030;  
 XX  
 DT 20-FEB-1998 (first entry)  
 XX  
 DE Antigenic human P53 peptide V10.  
 XX  
 KW Antigenic peptide; human papillomavirus; human P53; V10; MAGE gene;  
 KW human immunodeficiency virus; cancer antigen; tyrosinase; signal protein;  
 KW anthrax lethal factor; Lf; toxin; cationic fusion peptide; translocation;  
 KW gene therapy; polycationic affinity handle; therapeutic protein; LFN.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO9723236-A1.  
 XX  
 PD 03-JUL-1997.

XX  
 PF 13-DEC-1996; 96WO-US20463.  
 XX  
 PR 07-JUN-1996; 96US-0019275.  
 PR 13-DEC-1995; 95US-0008518.  
 XX  
 PA (HARD ) HARVARD COLLEGE.  
 XX  
 PI Ballard JD, Blanke SR, Collier RJ, Lyszak EL, Milne JC;  
 PI Scarnbach MN;  
 XX  
 DR WPI: 1997-350782/32.  
 XX  
 PT Introducing therapeutic proteins, especially antigens, into cells -  
 XX  
 PS using toxin molecules and/or polycationic handles for delivery  
 XX  
 XX Claim 15; Page 36; 67pp; English.  
 CC This is the antigenic human P53 peptide V10. This antigenic compound  
 CC can be introduced into the cytoplasm of a cell by a new method where  
 CC the cell is contacted with a fusion molecule comprising a delivery  
 CC molecule. The delivery molecule can either be a polycationic affinity  
 CC handle, LFN (the protective antigen binding domain of anthrax lethal  
 CC factor) or a toxin delivery molecule related to LFN. The antigenic  
 CC compound is linked to either of the delivery molecules by a covalent  
 CC bond. The B moiety of a toxin enhances delivery of the antigenic compound  
 CC into a cell. The anthrax toxin system of the invention eliminates the  
 CC need to generate fusion proteins with a toxin B moiety, which alleviates  
 CC problems associated with incorrect folding of lengthy fusion proteins.  
 CC Small cationic fusion peptides substituted for LFN may reduce the  
 CC possibility of steric interference with the biological activity of the  
 CC translocated protein. The method is used for the introduction of  
 CC antigens, e.g. MHC class I antigens or any other therapeutic protein,  
 CC e.g. toxin molecules, apoptosis-inducing molecules or signalling  
 CC proteins into the cells.  
 CC  
 XX  
 SQ Sequence 10 AA;  
 Query Match 82.9%; Score 29; DB 18; Length 10;  
 Best Local Similarity 85.7%; Pred. No. 2.5;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 FYQLALT 7  
 | | | | |  
 Db 1 FYQLALT 7  
 RESULT 8  
 AAY25186  
 ID AAY25186 standard; peptide: 14 AA.  
 XX  
 AC AAY25186;  
 XX  
 DT 03-SEP-1999 (first entry)  
 XX  
 DE Mutant P53 tumour suppressor protein peptide #2.  
 XX  
 KW Heat shock protein; HSP; complex; denatured protein matrix; antigen;  
 KW vaccine; allergic disease; treatment; susceptibility; TN2; skin rash;  
 KW allergic reaction; asthma; P53; tumour suppressor protein; mutant.  
 XX  
 OS Synthetic.  
 XX  
 PN MO9929182-A1.  
 XX  
 PD 17-JUN-1999.  
 XX  
 PF 04-DEC-1998; 98WO-US25734.  
 XX  
 PR 05-DEC-1997; 97US-0986234.  
 PR 05-DEC-1997; 97US-0985348.  
 XX



CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing  
CC inactive proteins or to supplement the patient's own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Pa,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH77196 to AAH7204  
CC and AAH7789 represent sequences used in the exemplification of the  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
SQ Sequence 64 AA:

Query Match 82.9%; Score 29; DB 22; Length 64;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 YQLALT 7  
|||  
Db 38 yqlalt 43

RESULT 11  
ID AA137291 standard; Protein: 491 AA.  
AA137291;  
07-OCT-1999 (first entry)  
Protein which is specific to Chlamydia trachomatis.  
Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
paratrachoma; inclusion conjunctivitis; genital disease; perlephleptitis;  
nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
batholinitis; pneumopathy; venereal lymphogranulomatosis.  
Chlamydia trachomatis.  
MO9928475-A2.  
10-JUN-1999.  
27-NOV-1998; 98MO-IB01939.  
04-NOV-1998; 98US-0107077.  
28-NOV-1997; 97FR-0015041.  
17-DEC-1997; 97FR-0016034.  
(GEST ) GENSET.  
Giffale R;  
WPI: 1999-371125/31.  
Genome sequence of Chlamydia trachomatis  
Disclosure; Page #VALUE1: 1755pp; English.  
AA13754-V37949 are encoded by open reading frames (ORFs) of the genome  
CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as  
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
CC can also be used to control growth of the microorganism. Chlamydia  
CC trachomatis is responsible for a large number of diseases, e.g. eye

CC diseases such as conventional trachoma, nonendemic trachoma,  
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
CC batholinitis, batholinitis; pneumopathy in breast feeding infants;  
CC and venereal lymphogranulomatosis. The polypeptides of the invention  
CC may be of use in treating these diseases.  
SQ Sequence 491 AA:

Query Match 82.9%; Score 29; DB 20; Length 491;  
Best Local Similarity 85.7%; Pred. No. 1,76+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYQLALT 7  
|||  
Db 49 fyqkalt 55

RESULT 12  
ID AAB85849 standard; Protein: 769 AA.  
AAB85849;  
29-OCT-2001 (first entry)  
Yeast Mlh1 protein.  
Hypemutable bacteria; mismatch repair gene; MMR gene; Muth; Muts; Muth;  
Muty; PMS2; Mlh1; Mlh3; PMSR; biocatalysis; bioremediation; biochemical;  
drug discovery; detoxification; toxin; biotransformation.  
Saccharomyces cerevisiae.  
MO200159092-A2.  
16-AUG-2001.  
12-FEB-2001; 2001MO-US04339.  
11-FEB-2000; 2000US-0181929.  
(UYJO ) UNIV JOHNS HOPKINS.  
Micolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;  
WPI: 2001-514664/56.  
N-PSDB; AAH76362.  
Making hypermutable bacteria for biocatalysis, bioremediation and drug  
discovery involves introducing polynucleotide comprising dominant  
negative allele of mismatch repair gene under regulatory sequence  
Example 1; Page 37-38; 68pp; English.  
The invention provides a method for generating a hypermutable bacteria.  
The method involves introducing a polynucleotide having a dominant  
negative allele of a mismatch repair (MMR) gene under the control of an  
inducible transcription regulatory sequence, into a bacterium. The cell  
becomes inducibly hypermutable. The method is useful to create desirable  
output traits for commercial applications, using dominant negative  
alleles of mismatch repair proteins. The mismatch repair gene is a Muth,  
Muts, Muth or Muty homologue and can be selected from PMS2, Mlh1, Mlh3,  
PMSR or PMSR homologue. The hypermutable bacteria is useful for the  
production, biocatalysis, bioremediation and drug discovery. It is also  
useful for detoxifying noxious chemicals from by-products of  
manufacturing processes or those used as catalysts for remediation of  
toxins present in the environment including polychlorobenzene, heavy  
metals and other environmental hazards for which there is a need to  
remove them from the environment. The hypermutable bacteria is also

CC useful for screening novel mutations in a gene or a set of genes that  
CC produce variant siblings that exhibit a new output trait not found in  
CC wild type cells. The bacteria are also useful for producing increased  
CC quantity or quality of protein or non-protein therapeutic molecule e.g.  
CC Penicillin G, Erythromycin and Clavulanic acid, by biotransformation.  
CC Dominant negative alleles of the MMR gene are useful for producing higher  
CC quantities of recombinant polypeptides. The present sequence represents  
CC a yeast MLH1 protein.  
XX  
XX  
SQ Sequence 769 AA:

Query Match 82.9%; Score 29; DB 22; Length 769;  
Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYQALAT 7  
|||: ||  
Db 570 FYQ191t 576

## RESULT 13

AAG63951  
ID AAG63951 standard; Protein; 769 AA.

AC AAG63951;

DT 29-OCT-2001 (first entry)

DE Amino acid sequence of yeast mismatch repair protein MLH1.

XX MLH1; mismatch repair gene; MMR gene; hypermutable yeast.

XX Unidentified.

XX WO200162945-A1.

XX 30-AUG-2001.

XX 21-FEB-2001; 2001WO-0505447.

XX 23-FEB-2000; 2000US-0184336.

XX (UYJO ) UNIV JOHNS HOPKINS.

XX (NICO/) NICOLAIDES N C.

XX (SASS/) SASS P M.

XX (GRASO/) GRASSO L.

XX (VOGE/) VOGELSTEIN B.

XX (KINZ/) KINZLER K W.

XX Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW.

XX WPI; 2001-522820/57.

XX N-PSDB; AAH75039.

XX Making hypermutable yeast that exhibit novel selected output traits for

XX commercial applications, comprises introducing polynucleotide

XX containing dominant negative allele of mismatch repair gene -

XX Disclosure; Page 36; 60pp; English.

XX The present sequence represents yeast MLH1. MLH1 is a mismatch repair  
XX (MMR) gene. The specification describes a method for making a  
XX hypermutable yeast, comprising introducing a polynucleotide containing  
XX a dominant negative allele of a mismatch repair (MMR) gene, into a  
XX yeast, whereby the cell becomes hypermutable. The method is useful  
XX to create desirable output traits for commercial applications, using  
XX dominant negative alleles of mismatch repair proteins. The hypermutable  
XX yeast is useful for production, biocatalysis, bioremediation and drug  
XX discovery. It is also useful in genetic screens for the direct selection  
XX of variant subclones that exhibit new output traits. The hypermutable  
XX yeast is also useful in the manufacturing industry for the generation  
XX of new biochemicals, for detoxifying noxious chemicals from by-products

CC of manufacturing processes or those used as catalysts, for remediation  
CC of toxins present in the environment including polychlorobenzene, heavy  
CC metals and other environmental hazards for which there is a need to  
CC remove them from the environment. The yeast is also useful for producing  
CC increased quantity or quality of protein or non-protein therapeutic  
CC molecule e.g., Penicillin G, Erythromycin and Clavulanic acid, by  
CC biotransformation.  
XX  
XX  
SQ Sequence 769 AA:

Query Match 82.9%; Score 29; DB 22; Length 769;  
Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYQALAT 7  
|||: ||  
Db 570 FYQ191t 576

## RESULT 14

AAR13166  
ID AAR13166 standard; Protein; 3391 AA.

AC AAR13166;

DT 21-NOV-1991 (first entry)

DE Proteins encoded by entire Dengue 2 virus genome.

XX dengue virus; detection; consensus sequence; Flavivirus; PCR.

XX Dengue virus.

XX Key Location/Qualifiers

XX Peptide 116..205

XX Protein 206..280 /label= PRM

XX Protein 281..775 /label= M

XX Protein 776..1127 /label= E

XX Protein 1128..1345 /label= NS1

XX Protein 1346..1474 /label= NS2A

XX Protein 1475..2093 /label= NS2B

XX Protein 2094..2243 /label= NS3

XX Protein 2244..2492 /label= NS4a

XX Protein 2493..3391 /label= NS4B

XX Protein 183 /label= NS5

XX Modified-site 347 /label= N-glycosylated

XX Modified-site 433 /label= N-glycosylated

XX Modified-site 905 /label= N-glycosylated

XX Modified-site 982 /label= N-glycosylated

XX Modified-site 1134 /label= N-glycosylated

XX Modified-site 1174 /label= N-glycosylated

XX Modified-site 1329 /label= N-glycosylated

XX Modified-site 1369 /label= N-glycosylated

```

FT Modified-site 2301
FT /label= N-glycosylated
FT Modified-site 2305
FT /label= N-glycosylated
FT Modified-site 2457
FT /label= N-glycosylated
FT Modified-site 2485
FT /label= N-glycosylated
FT Modified-site 2644
FT /label= N-glycosylated
FT Modified-site 2665
FT /label= N-glycosylated
FT Modified-site 2704
FT /label= N-glycosylated
FT Modified-site 2714
FT /label= N-glycosylated
FT
FT FR2654113-A.
FT
FT 10-MAY-1991.
FT
FT 09-NOV-1989; 89FR-0914724.
FT
FT 09-NOV-1989; 89FR-0014724.
FT
FT (INSP ) INST PASTEUR.
FT
FT Vincent D;
FT
FT WPI: 1991-225002/31.
FT N-PSDB; AAQ12787.
FT
FT Detection and identification of Flaviviridae in biological sample
FT - by amplifying consensus sequence then hybridisation opt.
FT followed by typing, e.g. sequencing amplified prod.
FT
FT
FT Disclosure; Fig 3; 24pp; French.
FT
FT
FT CC The dengue 2 virus is an example of a member of the Flaviviridae
FT CC which can be identified using the probe pair of the invention. A
FT CC species-specific sequence can be amplified using the claimed
FT CC oligonucleotides as primers in a PCR reaction (see AAQ12788 and
FT CC AAQ12789). Other viruses which can be identified include Japanese
FT CC encephalitis virus and yellow fever virus. All the dengue 2 virus
FT CC proteins are encoded from an uninterrupted genomic sequence.
FT
FT
FT SQ Sequence 3391 AA;

```

```

Query Match 82.9%; Score 29; DB 12; Length 3391;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2 YOLALT 7
Db 1270 yqialt 1275

```

```

RESULT 15
AAV37489
ID AAV37489 standard; Protein; 247 AA.
XX
XX AAV37489;
XX
XX 07-OCT-1999 (first entry)
XX
XX Chlamydia trachomatis transport protein.
XX
XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
XX paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
XX nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
XX bartolinitis; pneumopathy; venereal lymphogranulomatosis.
XX

```

```

OS Chlamydia trachomatis.
XX
XX W09928475-A2.
XX
XX 10-JUN-1999.
XX
XX 27-NOV-1998; 98WO-1B01939.
XX
XX 04-NOV-1998; 98US-0107077.
XX 28-NOV-1997; 97FR-0015041.
XX 17-DEC-1997; 97ER-0016034.
XX
XX (GEST ) GENSET.
XX
XX Griffiths R;
XX
XX WPI: 1999-371125/31.
XX
XX Genome sequence of Chlamydia trachomatis
XX
XX Disclosure; Page 1170; 1755pp; English.
XX
XX AAV36754-Y37949 are encoded by open reading frames (ORFs) of the genome
XX of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
XX vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
XX can also be used to control growth of the microorganism. Chlamydia
XX trachomatis is responsible for a large number of diseases, e.g. eye
XX diseases such as conventional trachoma, nonendemic trachoma,
XX paratrachoma, and inclusion conjunctivitis; genital diseases such as
XX nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
XX perihepatitis, bartolinitis; pneumopathy in breast feeding infants;
XX and venereal lymphogranulomatosis. The polypeptides of the invention
XX may be of use in treating these diseases.
XX
XX SQ Sequence 247 AA;

```

```

Query Match 80.0%; Score 28; DB 20; Length 247;
Best Local Similarity 71.4%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 FYOLALT 7
Db 167 fysialt 173

```

```

Search completed: July 15, 2002, 12:57:56
Job time: 412 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 12:59:17 ; Search time 75.67 seconds  
(without alignments)  
2.260 Million cell updates/sec

Title: US-09-712-819A-3  
Perfect score: 35  
Sequence: 1 FYQALMT 7

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : Issued Patents-AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	85.7	502	4	US-09-413-814-106
2	29	82.9	9	1	US-08-338-634-30
3	29	82.9	10	2	US-08-528-129A-5
4	29	82.9	12	2	US-08-528-129A-3
5	29	82.9	14	2	US-08-986-234-10
6	29	82.9	14	2	US-08-528-129A-4
7	29	82.9	21	2	US-08-528-129A-1
8	29	82.9	35	2	US-08-986-234-9
9	29	82.9	770	2	US-08-209-521-13
10	29	82.9	770	4	US-08-961-810-123
11	29	82.9	770	4	US-08-352-902D-123
12	27	77.1	278	2	US-07-857-224B-32
13	27	77.1	303	2	US-08-306-511A-9
14	27	77.1	303	2	US-08-893-274-9
15	27	77.1	303	3	US-08-581-918A-9
16	27	77.1	303	4	US-08-346-147B-9
17	27	77.1	303	4	US-08-822-936-9
18	27	77.1	303	5	PCR-US95-04636-9
19	27	77.1	555	3	US-08-501-573-3
20	27	77.1	555	3	US-09-040-444-3
21	27	77.1	618	2	US-08-770-761A-3
22	27	77.1	647	2	US-08-770-761A-8
23	27	77.1	660	2	US-08-770-761A-2
24	27	77.1	662	2	US-08-770-761A-5
25	27	77.1	705	2	US-08-770-761A-7
26	26	74.3	9	2	US-08-986-234-11
27	26	74.3	363	6	5223606-6

28	26	74.3	765	4	US-08-444-818-70	Sequence 70, Appl
29	26	74.3	933	4	US-08-764-870-14	Sequence 14, Appl
30	26	74.3	933	4	US-08-980-115-14	Sequence 14, Appl
31	25	71.4	65	1	US-08-035-917-14	Sequence 14, Appl
32	25	71.4	65	1	US-08-035-917-16	Sequence 16, Appl
33	25	71.4	65	1	US-08-055-917-18	Sequence 18, Appl
34	25	71.4	65	1	US-08-095-068-14	Sequence 14, Appl
35	25	71.4	65	1	US-08-095-068-16	Sequence 16, Appl
36	25	71.4	65	1	US-08-095-068-18	Sequence 18, Appl
37	25	71.4	65	1	US-08-140-721A-14	Sequence 14, Appl
38	25	71.4	65	1	US-08-140-721A-16	Sequence 16, Appl
39	25	71.4	65	1	US-08-140-721A-18	Sequence 18, Appl
40	25	71.4	65	1	US-08-619-790C-14	Sequence 14, Appl
41	25	71.4	65	1	US-08-619-790C-16	Sequence 16, Appl
42	25	71.4	65	1	US-08-619-790C-18	Sequence 18, Appl
43	25	71.4	65	2	US-07-785-565A-14	Sequence 14, Appl
44	25	71.4	65	2	US-07-785-565A-16	Sequence 16, Appl
45	25	71.4	65	2	US-07-785-565A-18	Sequence 18, Appl

## ALIGNMENTS

```
RESULT 1
US-09-413-814-106
Sequence 106, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Beyer, Stefan
APPLICANT: Bloeker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hottle, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
EARLIER FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 107
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 106
LENGTH: 502
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-106

Query Match 85.7%; Score 30; DB 4; Length 502;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FYQALMT 6
DB 20 FYQALMT 25

RESULT 2
US-08-338-634-30
Sequence 30, Application US/08338634
Patent No. 5679641
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Peptides of human p53 protein for use
in human T cell response inducing compositions, and
TITLE OF INVENTION: human p53 protein-specific cytotoxic T-lymphocytes.
```

NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann & Baron  
STREET: 350 Jericho Turnpike  
CITY: Jericho  
STATE: New York  
COUNTRY: United States of America  
ZIP: 11758  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/338,634  
FILING DATE: 06-February-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/NL93/00102  
FILING DATE: 18-May-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Baron, Ronald J.  
REGISTRATION NUMBER: 29,281  
REFERENCE/DOCKET NUMBER: 294-26  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (516) 822-3550  
TELEFAX: (516) 822-3582  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
US-08-338-634-30

Query Match 82.9%; Score 29; DB 1; Length 9;  
Best Local Similarity 85.7%; Pred. No. 1.7e+05;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FYOLAKT 7  
Db 3 FYOLAKT 9

RESULT 3  
US-08-528-129A-5  
Sequence 5, Application US/08528129A  
Patent No. 5997869  
GENERAL INFORMATION:  
APPLICANT: Goletz, Theresa J.  
APPLICANT: Berzofsky, Jay A.  
APPLICANT: Helman, Lee J.  
TITLE OF INVENTION: PEPTIDES CONTAINING A FUSION JOINT OF A  
TITLE OF INVENTION: CHIMERIC PROTEIN ENCODED BY DNA SPANNING A TUMOR-ASSOCIATED CH  
TITLE OF INVENTION: TRANSLOCATION AND THEIR USE AS IMMUNOGENS (as amended)  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSD for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/528,129A  
FILING DATE: 14-SEP-1995

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/424,573  
FILING DATE: 17-APR-1995  
APPLICATION NUMBER: 08/031,494  
FILING DATE: 15-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Wetherell, Jr., Ph.D. John R.  
REGISTRATION NUMBER: 31,678  
REFERENCE/DOCKET NUMBER: 08830/012001  
REFERENCE/DOCKET NUMBER: DHS Ref. No. 5997869 E-220-95/0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5039  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-528-129A-5

Query Match 82.9%; Score 29; DB 2; Length 10;  
Best Local Similarity 85.7%; Pred. No. 0.39;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FYOLAKT 7  
Db 1 FYOLAKT 7

RESULT 4  
US-08-528-129A-3  
Sequence 3, Application US/08528129A  
Patent No. 5997869  
GENERAL INFORMATION:  
APPLICANT: Goletz, Theresa J.  
APPLICANT: Berzofsky, Jay A.  
APPLICANT: Helman, Lee J.  
TITLE OF INVENTION: PEPTIDES CONTAINING A FUSION JOINT OF A  
TITLE OF INVENTION: CHIMERIC PROTEIN ENCODED BY DNA SPANNING A TUMOR-ASSOCIATED  
TITLE OF INVENTION: TRANSLOCATION AND THEIR USE AS IMMUNOGENS (as amended)  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSD for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/528,129A  
FILING DATE: 14-SEP-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/424,573  
FILING DATE: 17-APR-1995  
APPLICATION NUMBER: 08/031,494  
FILING DATE: 15-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Wetherell, Jr., Ph.D. John R.  
REGISTRATION NUMBER: 31,678  
REFERENCE/DOCKET NUMBER: 08830/012001  
REFERENCE/DOCKET NUMBER: DHS Ref. No. 5997869 E-220-95/0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5039

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-528-129A-3

Query Match 82.9%; Score 29; DB 2; Length 12;  
Best Local Similarity 85.7%; Pred. No. 0.48;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYOLAKT 7  
|||||  
DB 6 FYOLAKT 12

RESULT 5  
US-08-986-234-10  
Sequence 10, Application US/08986234  
Patent No. 5981706

GENERAL INFORMATION:

APPLICANT: Hallen, et al.

TITLE OF INVENTION: Methods for Synthesizing Heat Shock Protein Complexes

FILE REFERENCE: UNME-0008-1

CURRENT APPLICATION NUMBER: US/08/986,234

CURRENT FILING DATE: 1997-12-05

NUMBER OF SEQ ID NOS: 114

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 10

LENGTH: 14

TYPE: PRT

ORGANISM: human

US-08-986-234-10

Query Match 82.9%; Score 29; DB 2; Length 14;  
Best Local Similarity 85.7%; Pred. No. 0.58;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYOLAKT 7  
|||||  
DB 3 FYOLAKT 9

RESULT 6  
US-08-528-129A-4

Sequence 4, Application US/08528129A

Patent No. 5997869

GENERAL INFORMATION:

APPLICANT: Goletz, Theresa J.

APPLICANT: Berzofsky, Jay A.

APPLICANT: Helman, Lee J.

TITLE OF INVENTION: PEPTIDES CONTAINING A FUSION JOINT OF A

TITLE OF INVENTION: CHIMERIC PROTEIN ENCODED BY DNA SPANNING A TUMOR-ASSOCIATED CH

TITLE OF INVENTION: TRANSLLOCATION AND THEIR USE AS IMMUNOGENS (as amended)

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: CA

COUNTRY: US

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/528,129A

FILING DATE: 14-SEP-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/424,573

FILING DATE: 17-APR-1995

APPLICATION NUMBER: 08/031,494

FILING DATE: 15-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Wetherell, Jr., Ph.D. John R.

REGISTRATION NUMBER: 31,678

REFERENCE/DOCKET NUMBER: 08830/012001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-678-5070

TELEFAX: 619-678-5099

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 14 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

Query Match 82.9%; Score 29; DB 2; Length 14;  
Best Local Similarity 85.7%; Pred. No. 0.58;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYOLAKT 7  
|||||  
DB 3 FYOLAKT 9

RESULT 7  
US-08-528-129A-1

Sequence 1, Application US/08528129A

Patent No. 5997869

GENERAL INFORMATION:

APPLICANT: Goletz, Theresa J.

APPLICANT: Berzofsky, Jay A.

APPLICANT: Helman, Lee J.

TITLE OF INVENTION: PEPTIDES CONTAINING A FUSION JOINT OF A

TITLE OF INVENTION: CHIMERIC PROTEIN ENCODED BY DNA SPANNING A TUMOR-ASSOCIATED

TITLE OF INVENTION: TRANSLLOCATION AND THEIR USE AS IMMUNOGENS (as amended)

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: CA

COUNTRY: US

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/528,129A

FILING DATE: 14-SEP-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/424,573

FILING DATE: 17-APR-1995

APPLICATION NUMBER: 08/031,494

FILING DATE: 15-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Wetherell, Jr., Ph.D. John R.

REGISTRATION NUMBER: 31,678

REFERENCE/DOCKET NUMBER: 08830/012001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-678-5070

TELEFAX: 619-678-5099

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-528-129A-1

Query Match  
Best Local Similarity 82.9%; Score 29; DB 2; Length 21;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FYQLAKT 7  
|||||  
DB 10 FYQLAKT 16

RESULT 8  
US-08-986-234-9  
Sequence 9, Application US/08986234  
Patent No. 5981706  
GENERAL INFORMATION:  
APPLICANT: Wallen, et al.  
TITLE OF INVENTION: Methods for synthesizing Heat Shock Protein Complexes  
FILE REFERENCE: UNME-0008-1  
CURRENT APPLICATION NUMBER: US/08/986,234  
CURRENT FILING DATE: 1997-12-05  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 9  
LENGTH: 35  
TYPE: PRT  
ORGANISM: human  
US-08-986-234-9

Query Match  
Best Local Similarity 82.9%; Score 29; DB 2; Length 35;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FYQLAKT 7  
|||||  
DB 10 FYQLAKT 16

RESULT 9  
US-08-209-521-13  
Sequence 13, Application US/08209521  
Patent No. 5922855  
GENERAL INFORMATION:  
APPLICANT: Liskay, Robert M.  
APPLICANT: Bronner, C. Eric  
APPLICANT: Baker, Sean M.  
APPLICANT: Bollag, Ron J.  
APPLICANT: Kolodner, Richard D.  
TITLE OF INVENTION: MAMMALIAN DNA MISMATCH REPAIR GENES  
TITLE OF INVENTION: hMLH1 AND hPMS1  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kolisch, Hartwell, Dickinson, McCormack &  
STREET: 520 S.W. Yamhill, Suite 200  
CITY: Portland  
STATE: Oregon  
COUNTRY: US  
ZIP: 97204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/209,521  
FILING DATE: 08-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Van Rysselberghe, Pierre C.  
REGISTRATION NUMBER: 33,557  
REFERENCE/DOCKET NUMBER: OHSU 306A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (503) 224-6655  
TELEFAX: (503) 224-6679  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-209-521-13

Query Match  
Best Local Similarity 82.9%; Score 29; DB 2; Length 770;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FYQLGKT 7  
|||||  
DB 571 FYQLGKT 577

RESULT 10  
US-08-961-810-123  
Sequence 123, Application US/08961810  
Patent No. 6165713  
GENERAL INFORMATION:  
APPLICANT: Liskay, Robert M.  
APPLICANT: Bronner, C. Eric  
APPLICANT: Baker, Sean M.  
APPLICANT: Bollag, Ron J.  
APPLICANT: Kolodner, Richard D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS RELATING TO DNA  
TITLE OF INVENTION: MISMATCH REPAIR GENES  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kolisch, Hartwell, Dickinson, McCormack &  
STREET: 520 S.W. Yamhill Street, Suite 200  
CITY: Portland  
STATE: Oregon  
COUNTRY: U.S.A.  
ZIP: 97204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,810  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Van Rysselberghe, Pierre C.  
REGISTRATION NUMBER: 33,557  
REFERENCE/DOCKET NUMBER: OHSU 306B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (503) 224-6655  
TELEFAX: (503) 295-6679  
INFORMATION FOR SEQ ID NO: 123:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-961-810-123

Query Match 82.9%; Score 29; DB 4; Length 770;  
Best Local Similarity 71.4%; Pred. No. 58;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYQIALT 7  
DB 571 FYQIALT 577

RESULT 11  
US-08-352-902D-123  
Sequence 123, Application US/08352902D  
Patent No. 6191268  
GENERAL INFORMATION:  
APPLICANT: Liskay, Robert M.  
Bromner, C. Eric  
Baker, Sean M.  
Bollag, Roni J.  
Kolodner, Richard D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS RELATING TO DNA  
NUMBER OF SEQUENCES: 149  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kolisch, Hartwell, Dickinson, McCormack &  
Heuser  
STREET: 520 S.W. Yamhill Street, Suite 200  
CITY: Portland  
STATE: Oregon  
COUNTRY: U.S.A.  
ZIP: 97204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/352,902D  
FILING DATE: 09-Dec-1994  
CLASSIFICATION: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Van Ryselberghe, Pierre C.  
REGISTRATION NUMBER: 33,557  
REFERENCE/DOCKET NUMBER: OHSU 306B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (503) 224-6555  
TELEFAX: (503) 295-6679  
TELEX: 360619  
INFORMATION FOR SEQ ID NO: 123:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 770 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 123:  
US-08-352-902D-123

Query Match 82.9%; Score 29; DB 4; Length 770;  
Best Local Similarity 71.4%; Pred. No. 58;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FYQIALT 7  
DB 571 FYQIALT 577

RESULT 12  
US-07-857-224B-32

Sequence 32, Application US/07857224B  
Patent No. 5958784  
GENERAL INFORMATION:  
APPLICANT: Benner, Steven A.  
TITLE OF INVENTION: Predicting Folded Structures of Proteins  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Steven A. Benner  
STREET: Hadlaubstrasse 151  
CITY: Zurich  
STATE: none  
COUNTRY: Switzerland  
ZIP: (note: this is an international post code) CH-8092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage  
OPERATING SYSTEM: Macintosh 7.0  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/857,224B  
FILING DATE: 03/25/92  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA: none  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (International) 41 1 632 2830  
TELEFAX: (International) 41 1 262 2437  
TELEX: none  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 278  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: human  
FEATURE: Protein kinase; Table 8 Column 36  
PUBLICATION INFORMATION:  
AUTHORS:  
AUTHORS: Hanks, S. K.  
AUTHORS: Quinn, A. M.  
AUTHORS: Hunter, T.  
TITLE: The protein kinase family  
JOURNAL: Science  
VOLUME: 241  
PAGES: 42-52  
DATE: 1988  
US-07-857-224B-32

Query Match 77.1%; Score 27; DB 2; Length 278;  
Best Local Similarity 83.3%; Pred. No. 55;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 YQIALT 7  
DB 164 YQIALT 169

RESULT 13  
US-08-306-511A-9  
Sequence 9, Application US/08306511A  
Patent No. 5962316  
GENERAL INFORMATION:  
APPLICANT: Beach, David H.  
APPLICANT: Demetrick, Douglas J.  
APPLICANT: Serrano, Manuel  
APPLICANT: Hannon, Gregory J.  
TITLE OF INVENTION: Cell-Cycle Regulatory Proteins, and Uses  
RELATED THERETO  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD

STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII(text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/306,511A  
FILING DATE: 14-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: CSI-001CP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: N-terminal  
US-08-306-511A-9

Query Match 77.1%; Score 27; DB 2; Length 303;  
Best Local Similarity 83.3%; Pred. No. 60;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 YOLALT 7  
11:111  
Db 167 YOMALT 172

RESULT 14  
US-08-893-274-9  
Sequence 9, Application US/08893274  
Patent No. 5968821  
GENERAL INFORMATION:  
APPLICANT: Beach, David H.  
APPLICANT: Demetrick, Douglas J.  
APPLICANT: Serrano, Manuel  
TITLE OF INVENTION: Cell-Cycle Regulatory Proteins,  
TITLE OF INVENTION: and Uses Related Thereto  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELLIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII(text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/893,274  
FILING DATE: 15-JULY-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/306,511  
FILING DATE: 14-SEPTEMBER-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/248,812  
FILING DATE: 25-MAY-1994

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,371  
FILING DATE: 14-APRIL-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/154,915  
FILING DATE: 18-NOVEMBER-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/991,997  
FILING DATE: 17-DECEMBER-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/963,308  
FILING DATE: 16-OCTOBER-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: MIV-071.09  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 832-7000  
TELEFAX: (617) 832-7000  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: N-terminal  
US-08-893-274-9

Query Match 77.1%; Score 27; DB 2; Length 303;  
Best Local Similarity 83.3%; Pred. No. 60;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 YOLALT 7  
11:111  
Db 167 YOMALT 172

RESULT 15  
US-08-581-918A-9  
Sequence 9, Application US/08581918A  
Patent No. 6043030  
GENERAL INFORMATION:  
APPLICANT: Beach, David H.  
APPLICANT: Demetrick, Douglas J.  
APPLICANT: Serrano, Manuel  
TITLE OF INVENTION: Cell-Cycle Regulatory Proteins, and Uses  
TITLE OF INVENTION: Related Thereto  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELLIOT  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Wordpad  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/581,918A  
FILING DATE: 02-JAN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/497,214  
FILING DATE: 30-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/346,147  
FILING DATE: 29-NOV-1994  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/306,511  
FILING DATE: 14-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/248,812  
FILING DATE: 25-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,371  
FILING DATE: 14-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/154,915  
FILING DATE: 18-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/991,997  
FILING DATE: 17-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: MIV-071.06  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 832-1299  
TELEFAX: (617) 832-7000  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: N-terminal  
US-08-581-918A-9

Query Match 77.1%; Score 27; DB 3; Length 303;  
Best Local Similarity 83.3%; Pred No. 60;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 YQALTY 7  
11:111  
Db 167 YQALTY 172

Search completed: July 15, 2002, 12:59:17  
Job time: 388 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:01:01 ; Search time 95.45 seconds  
(without alignments)  
8.054 Million cell updates/sec

Title: US-09-712-819A-4  
Perfect score: 46  
Sequence: 1 PKLIIYWA 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	46	100.0	81	2	PH1048
2	46	100.0	92	2	S37533
3	46	100.0	92	2	S37530
4	46	100.0	92	2	S37529
5	46	100.0	92	2	S37535
6	46	100.0	92	2	S37534
7	46	100.0	92	2	S37532
8	46	100.0	96	2	G38601
9	46	100.0	98	2	S19974
10	46	100.0	101	2	S26337
11	46	100.0	101	2	PH1046
12	46	100.0	101	2	PH0869
13	46	100.0	103	2	PH1050
14	46	100.0	103	2	PH1051
15	46	100.0	103	2	PH1052
16	46	100.0	103	2	PH1054
17	46	100.0	103	2	PH1047
18	46	100.0	104	2	PH1101
19	46	100.0	104	2	PH1102
20	46	100.0	104	2	PH1103
21	46	100.0	104	2	PH1104
22	46	100.0	104	2	F38601
23	46	100.0	107	2	S36268
24	46	100.0	111	2	S03304
25	46	100.0	111	2	G30502
26	46	100.0	112	2	PL0265
27	46	100.0	112	2	S09970
28	46	100.0	112	2	S41393
29	46	100.0	112	2	S43103

30	46	100.0	113	2	PR0407	Ig light chain V r
31	46	100.0	113	2	PR0408	Ig light chain V r
32	46	100.0	113	2	S30523	Ig kappa chain V r
33	46	100.0	113	2	S30520	Ig kappa chain V r
34	46	100.0	113	2	PL0263	Ig kappa chain V r
35	46	100.0	113	2	A49260	antitumor monoclon
36	46	100.0	113	2	S34003	Ig kappa chain V r
37	46	100.0	113	2	S34002	Ig kappa chain V r
38	46	100.0	114	1	K4HUN	Ig kappa chain V-I
39	46	100.0	114	2	S44119	Ig kappa chain V r
40	46	100.0	117	2	S42466	Ig kappa chain V r
41	46	100.0	118	2	PT0356	Ig kappa chain V r
42	46	100.0	120	2	G33932	Ig kappa chain pre
43	46	100.0	120	2	S51147	antibody light cha
44	46	100.0	121	1	K4HU	Ig kappa chain pre
45	46	100.0	131	2	PL0207	anti-idiotypic ant

## ALIGNMENTS

RESULT 1  
PH1048  
Ig light chain V region (clone 165.49) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 15-Jun-1996  
R:Accession: PH1048  
R:Tillman, D.M.; Jou, N.T.; Hill, R.J.; Marion, T.N.  
J. Exp. Med. 176, 761-779, 1992  
A:Title: Both IgM and IgG anti-DNA antibodies are the products of clonally selective  
A:Reference number: PH0971; MUID:92381444  
A:Accession: PH1048  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-81 <TIL>  
A:Experimental source: B cell, strain [NZB x NZM]F1  
A:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: Immunoglobulin

Query Match 100.0%; Score 46; DB 2; Length 81;  
Best Local Similarity 100.0%; Pred. No. 0.067;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PKLIIYWA 8  
DB 29 PKLIIYWA 36  
RESULT 2  
S37533  
Ig kappa chain V region (V-kappa 4) - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999  
C:Accession: S37533  
R:Klein, U.; Kuipers, R.; Rajewsky, K.  
Submitted to the EMBL Data Library, September 1993  
A:Description: Human IgM(+)IgD(+) cells, the major B cell subset in the peripheral bl  
A:Reference number: S37501  
A:Accession: S37533  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-92 <KLE>  
A:Cross-references: EMBL:226627; NID:9405708; PIDN:CAA81380.1; PID:9405709  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotrimer; Immunoglobulin

Query Match 100.0%; Score 46; DB 2; Length 92;  
Best Local Similarity 100.0%; Pred. No. 0.075;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PKLIIYWA 8

Db 28 PKLITYWA 35

## RESULT 3

Ig kappa chain V region (V-kappa 4) - human (fragment)  
S37530  
C:Species: Homo sapiens (man)  
C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999  
C:Accession: S37530  
R:Klein, U.; Kuipers, R.; Rajewsky, K.  
submitted to the EMBL Data Library, September 1993  
A:Description: Human IGM(+)19D(+) cells, the major B cell subset in the peripheral blood  
A:Reference number: S37501  
A:Accession: S37530  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-92 <KLE>  
A:Cross-references: EMBL:226630; NID:g405702; PIDN:CAA81384.1; PID:g405703  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin

Query Match 100.0%; Score 46; DB 2; Length 92;  
Best Local Similarity 100.0%; Pred. No. 0.075;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PKLITYWA 8  
Db 28 PKLITYWA 35

## RESULT 4

Ig kappa chain V region (V-kappa 4) - human (fragment)  
S37529  
C:Species: Homo sapiens (man)  
C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999  
C:Accession: S37529  
R:Klein, U.; Kuipers, R.; Rajewsky, K.  
submitted to the EMBL Data Library, September 1993  
A:Description: Human IGM(+)19D(+) cells, the major B cell subset in the peripheral blood  
A:Reference number: S37529  
A:Accession: S37529  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-92 <KLE>  
A:Cross-references: EMBL:226631; NID:g405700; PIDN:CAA81384.1; PID:g405701  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin

Query Match 100.0%; Score 46; DB 2; Length 92;  
Best Local Similarity 100.0%; Pred. No. 0.075;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PKLITYWA 8  
Db 28 PKLITYWA 35

## RESULT 5

Ig kappa chain V region (V-kappa 4) - human (fragment)  
S37535  
C:Species: Homo sapiens (man)  
C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999  
C:Accession: S37535  
R:Klein, U.; Kuipers, R.; Rajewsky, K.  
submitted to the EMBL Data Library, September 1993  
A:Description: Human IGM(+)19D(+) cells, the major B cell subset in the peripheral blood  
A:Reference number: S37501  
A:Accession: S37535  
A:Status: preliminary  
A:Molecule type: mRNA

A:Residues: 1-92 <KLE>  
A:Cross-references: EMBL:226625; NID:g405712; PIDN:CAA81378.1; PID:g405713  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin

Query Match 100.0%; Score 46; DB 2; Length 92;  
Best Local Similarity 100.0%; Pred. No. 0.075;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PKLITYWA 8  
Db 28 PKLITYWA 35

## RESULT 6

Ig kappa chain V region (V-kappa 4) - human (fragment)  
S37534  
C:Species: Homo sapiens (man)  
C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999  
C:Accession: S37534  
R:Klein, U.; Kuipers, R.; Rajewsky, K.  
submitted to the EMBL Data Library, September 1993  
A:Description: Human IGM(+)19D(+) cells, the major B cell subset in the peripheral blood  
A:Reference number: S37501  
A:Accession: S37534  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-92 <KLE>  
A:Cross-references: EMBL:226626; NID:g405710; PIDN:CAA81379.1; PID:g405711  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin

Query Match 100.0%; Score 46; DB 2; Length 92;  
Best Local Similarity 100.0%; Pred. No. 0.075;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PKLITYWA 8  
Db 28 PKLITYWA 35

## RESULT 7

Ig kappa chain V region (V-kappa 4) - human (fragment)  
S37532  
C:Species: Homo sapiens (man)  
C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999  
C:Accession: S37532  
R:Klein, U.; Kuipers, R.; Rajewsky, K.  
submitted to the EMBL Data Library, September 1993  
A:Description: Human IGM(+)19D(+) cells, the major B cell subset in the peripheral blood  
A:Reference number: S37501  
A:Accession: S37532  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-92 <KLE>  
A:Cross-references: EMBL:226628; NID:g405706; PIDN:CAA81381.1; PID:g405707  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin

Query Match 100.0%; Score 46; DB 2; Length 92;  
Best Local Similarity 100.0%; Pred. No. 0.075;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PKLITYWA 8  
Db 28 PKLITYWA 35

## RESULT 8

G38601

Ig kappa chain V region (4A9) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Aug-1991 #sequence\_revision 30-Aug-1991 #text\_change 23-Jul-1999  
C:Accession: G38601  
R:Goshorn, S.C.; Retzel, E.; Jemerson, R.  
J. Biol. Chem. 266, 2134-2142, 1991  
A:Title: Common structural features among monoclonal antibodies binding the same antigen  
A:Reference number: A38601; MUID:91115823  
A:Accession: G38601  
A>Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-96 <GOS>  
A:Cross-references: GB:M57984; NID:g196414; PIDN:AAA63365.1; PID:g196415  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 46; DB 2; Length 96;  
Best Local Similarity 100.0%; Pred. No. 0.078;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLIYWA 8  
|||||||  
DB 32 PKLIYWA 39

RESULT 9  
S19974  
Ig kappa chain V region (M-T406) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 21-Jan-2000  
C:Accession: S19974  
R:Meisenhorn, W.; Rietmuller, G.; Weiss, E.M.; Rieber, E.P.  
submitted to the EMBL Data Library March 1992  
A:Description: Structural characterization of CD4 mAb.  
A:Reference number: S19963  
A:Accession: S19974  
A>Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-98 <MEI>  
A:Cross-references: EMBL:X65096; NID:g52294; PIDN:CAA6224.1; PID:g52295  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:1-80/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 46; DB 2; Length 98;  
Best Local Similarity 100.0%; Pred. No. 0.08;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLIYWA 8  
|||||||  
DB 34 PKLIYWA 41

RESULT 10  
S26337  
Ig light chain V region - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 19-Mar-1998 #sequence\_revision 19-Mar-1998 #text\_change 20-Jun-2000  
C:Accession: S26337; S78449  
R:Stark, S.E.; Caton, A.J.  
J. Exp. Med. 174, 613-624, 1991  
A:Title: Antibodies that are specific for a single amino acid interchange in a protein  
A:Reference number: S26309; MUID:91341421  
A:Accession: S26337  
A:Molecule type: mRNA  
A:Residues: 1-101 <STA>  
A:Cross-references: EMBL:X59193  
R:Caton, A.J.  
submitted to the EMBL Data Library, April 1991  
A:Reference number: S78447  
A:Accession: S78449

A:Molecule type: mRNA  
A:Residues: 1-60, 'T', 62-91, 'S', 93-101 <CAT>  
A:Cross-references: EMBL:X59193; NID:g52323; PIDN:CAA41903.1; PID:g1334067  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:8-88/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 46; DB 2; Length 101;  
Best Local Similarity 100.0%; Pred. No. 0.082;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLIYWA 8  
|||||||  
DB 42 PKLIYWA 49

RESULT 11  
PH1046  
Ig light chain V region (clone 202.9) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 21-Jan-2000  
C:Accession: PH1046  
R:Tillman, D.M.; Jou, N.T.; Hill, R.J.; Marlon, T.N.  
J. Exp. Med. 176, 761-779, 1992  
A:Title: Both IgM and IgG anti-DNA antibodies are the products of clonally selective  
A:Reference number: PH0971; MUID:92381444  
A:Accession: PH1046  
A>Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-101 <TIL>  
A:Experimental source: B cell, strain [NZB x NZW]F1  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: immunoglobulin  
F:16-96/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 46; DB 2; Length 101;  
Best Local Similarity 100.0%; Pred. No. 0.082;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLIYWA 8  
|||||||  
DB 50 PKLIYWA 57

RESULT 12  
PH0869  
Ig kappa chain V region (anti-DNA, H2F) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 08-Oct-1992 #sequence\_revision 09-Oct-1992 #text\_change 21-Jan-2000  
C:Accession: PH0869  
R:Manheimer-Lory, A.; Katz, J.B.; Pillingner, M.; Ghossein, C.; Smith, A.; Diamond, B.  
J. Exp. Med. 174, 1639-1652, 1991  
A:Title: Molecular characteristics of antibodies bearing an anti-DNA-associated Idiot.  
A:Reference number: PH0862; MUID:92078875  
A:Accession: PH0869  
A:Molecule type: DNA  
A:Residues: 1-101 <MAN>  
A:Note: residues 28-33 were obtained from Figure 4  
C:Comment: This antibody is produced by Epstein-Barr virus-transformed B cell that be  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:1-23/Region: Framework 1  
F:16-96/Domain: Immunoglobulin homology <IMM>  
F:24-40/Region: complementarity-determining 1  
F:41-55/Region: complementarity-determining 2  
F:56-62/Region: complementarity-determining 3  
F:63-94/Region: framework 3  
F:95-101/Region: complementarity-determining 3

Query Match 100.0%; Score 46; DB 2; Length 101;

Best Local Similarity 100.0%; Pred. No. 0.082;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PKLLIYWA 8  
|||||||  
Db 50 PKLLIYWA 57

## RESULT 13

PH1050  
Ig light chain V region (clone 111-cl) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 21-Jan-2000  
C:Accession: PH1050  
R:Tillman, D.M.; Jou, N.T.; Hill, R.J.; Marion, T.N.  
J. Exp. Med. 176, 761-779, 1992  
A:Title: Both IgM and IgG anti-DNA antibodies are the products of clonally selective B  
A:Reference number: PH0971; M01D:92381444  
A:Accession: PH1050  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-103 <TIL>  
A:Experimental source: B cell, strain [NZB x NZW]F1  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: Immunoglobulin  
F:16-96/Domain: Immunoglobulin homology <IMM>

## Query Match

Best Local Similarity 100.0%; Score 46; DB 2; Length 103;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLLIYWA 8  
|||||||  
Db 50 PKLLIYWA 57

## RESULT 14

PH1051  
Ig light chain V region (clone 165.3) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 21-Jan-2000  
C:Accession: PH1051  
R:Tillman, D.M.; Jou, N.T.; Hill, R.J.; Marion, T.N.  
J. Exp. Med. 176, 761-779, 1992  
A:Title: Both IgM and IgG anti-DNA antibodies are the products of clonally selective B  
A:Reference number: PH0971; M01D:92381444  
A:Accession: PH1051  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-103 <TIL>  
A:Experimental source: B cell, strain [NZB x NZW]F1  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: Immunoglobulin  
F:16-96/Domain: Immunoglobulin homology <IMM>

## Query Match

Best Local Similarity 100.0%; Score 46; DB 2; Length 103;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLLIYWA 8  
|||||||  
Db 50 PKLLIYWA 57

## RESULT 15

PH1052  
Ig light chain V region (clone 165.5) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 21-Jan-2000  
C:Accession: PH1052  
R:Tillman, D.M.; Jou, N.T.; Hill, R.J.; Marion, T.N.

J. Exp. Med. 176, 761-779, 1992  
A:Title: Both IgM and IgG anti-DNA antibodies are the products of clonally selective  
A:Reference number: PH0971; M01D:92381444  
A:Accession: PH1052  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-103 <TIL>  
A:Experimental source: B cell, strain [NZB x NZW]F1  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: Immunoglobulin  
F:16-96/Domain: Immunoglobulin homology <IMM>

## Query Match

Best Local Similarity 100.0%; Score 46; DB 2; Length 103;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLLIYWA 8  
|||||||  
Db 50 PKLLIYWA 57

Search completed: July 15, 2002, 13:01:01  
Job time: 467 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:23:32 ; Search time 44.9 Seconds  
(without alignments)  
6.899 Million cell updates/sec

Title: US-09-712-819A-4

Perfect score: 46  
Sequence: 1 PKLIYNA 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query	Match length	ID	Description
	1	46	100.0	114	1	KV4A_HUMAN
	2	46	100.0	121	1	KV4Q_HUMAN
	3	46	100.0	133	1	KV4B_HUMAN
	4	46	100.0	134	1	KV4C_HUMAN
	5	35	76.1	151	1	YCRC_BACSU
	6	35	76.1	283	1	YORS_ADEG1
	7	35	76.1	705	1	PURL_PYRHO
	8	34	73.9	146	1	AP4A_HUMAN
	9	34	73.9	146	1	AP4A_MOUSE
	10	34	73.9	146	1	AP4A_PIG
	11	34	73.9	3224	1	RBP2_HUMAN
	12	33	71.7	108	1	KV1M_HUMAN
	13	33	71.7	108	1	KV5Q_MOUSE
	14	33	71.7	108	1	KV5R_MOUSE
	15	33	71.7	108	1	KV5S_MOUSE
	16	33	71.7	108	1	KV5T_MOUSE
	17	33	71.7	111	1	KV3E_MOUSE
	18	33	71.7	111	1	KV3I_MOUSE
	19	33	71.7	111	1	KV3K_MOUSE
	20	33	71.7	111	1	KV3O_MOUSE
	21	33	71.7	111	1	KV3R_MOUSE
	22	33	71.7	111	1	KV3S_MOUSE
	23	33	71.7	111	1	KV3J_MOUSE
	24	33	71.7	111	1	KV3U_MOUSE
	25	33	71.7	114	1	KV1A_MOUSE
	26	33	71.7	131	1	KV3I_MOUSE
	27	33	71.7	136	1	KV5H_MOUSE
	28	33	71.7	331	1	HUGA_DOLMA
	29	33	71.7	331	1	HUGA_VESUV
	30	33	71.7	1339	1	DPOA_TRYRB
	31	32	69.6	94	1	KV1I_RABIT
	32	32	69.6	108	1	KV0S_RABIT
	33	32	69.6	108	1	KV0C_RABIT

34	32	69.6	108	1	KV0T_RABIT	P01688 oryctolagus
35	32	69.6	108	1	KV0K_RABIT	P01689 oryctolagus
36	32	69.6	108	1	KV1K_HUMAN	P01603 homo sapien
37	32	69.6	108	1	KV1O_HUMAN	P01604 homo sapien
38	32	69.6	108	1	KV1Q_HUMAN	P01607 homo sapien
39	32	69.6	108	1	KV1S_HUMAN	P01611 homo sapien
40	32	69.6	109	1	KV1A_RABIT	P01695 oryctolagus
41	32	69.6	110	1	KV13_RABIT	P01694 oryctolagus
42	32	69.6	110	1	KV15_RABIT	P01696 oryctolagus
43	32	69.6	110	1	KV3P_MOUSE	P01668 mus musculu
44	32	69.6	111	1	KV12_RABIT	P01693 oryctolagus
45	32	69.6	111	1	KV3A_MOUSE	P01654 mus musculu

## ALIGNMENTS

```
RESULT 1
KV4A_HUMAN STANDARD; PRT: 114 AA.
ID KV4A_HUMAN
AC P01625;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-IV region Len.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=76004342; PubMed=50995;
RX Schneider M., Hirschmann N.;
RA "The primary structure of a monoclonic immunoglobulin-L-chain of
RT subgroup IV of the kappa type (Bence-Jones protein Len).";
RL Hoppe-Seyler's Z. Physiol. Chem. 356:357(1975).
RN [2]
RP REVISION TO 9.
RA Salomon A.;
RL Submitted (AUG-1996) to the SWISS-PROT data bank.
CC -!- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (3) MARKER.
CC PIR: A01903; K4H01N.
DR HSSP; P01789; IMCP.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00406; IGV_1.
KW Immunoglobulin V region; Bence-Jones protein.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 40 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 41 55 FRAMEWORK-2.
FT DOMAIN 56 62 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 63 94 FRAMEWORK-3.
FT DOMAIN 95 101 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 102 113 FRAMEWORK-4.
FT DISULFID 23 94 BY SIMILARITY.
FT NON_TER 114 114
SQ SEQUENCE 114 AA; 12640 MW; 0647F1D17F236485 CRC64;
```

Query Match 100.0%; Score 46; DB 1; Length 114;  
Best Local Similarity 100.0%; Pred. No. 0.03;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKLIYNA 8  
Db 50 PKLIYNA 57

RESULT 2  
KV4Q\_HUMAN STANDARD; PRT: 121 AA.

```

AC P06312;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-IV region precursor (FragmentL).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=86041853; PubMed=2997712;
RA Klobbeck H.G., Bornkamm G.W., Combariato G., Mochkat R., Pohlenz H.D.,
RA Zachau H.G.;
RT "Subgroup IV of human Immunoglobulin K light chains is encoded by a
RT single germline gene."
RL Nucleic Acids Res. 13:6515-6529(1985).
CC -----
CC -1- MISCELLANEOUS: THERE IS ONLY ONE IG KAPPA V-IV GENE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Z00023; CAA77318.1; -
CC PIR: A01902; KAHU.
CC HSSP: P01789; 1MCP.
CC InterPro: IPR003006; IG_MHC.
CC InterPro: IPR003596; IG_V.
CC Pfam: PF00047; 1g; 1.
CC SMART: SM00406; 1g; 1.
CC Immunoglobulin V region; Signal.
KW SIGNAL
FT CHAIN 1 20 IG KAPPA CHAIN V-IV REGION.
FT DOMAIN 21 >121 FRAMEWORK-1.
FT DOMAIN 44 60 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 61 75 FRAMEWORK-2.
FT DOMAIN 76 82 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 83 114 FRAMEWORK-3.
FT DOMAIN 115 121 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 43 114 BY SIMILARITY.
FT NON_TER 121 121
SQ SEQUENCE 121 AA; 13380 MW; 9586AD418803974 CRC64;

Query Match 100.0%; Score 46; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 0.032;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLLIYMA 8
Db 70 PKLLIYMA 77

RESULT 3
KV4B_HUMAN STANDARD; PRT; 133 AA.
AC P06313;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-IV region J1 precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=86041853; PubMed=2997712;
RA Klobbeck H.G., Bornkamm G.W., Combariato G., Mochkat R., Pohlenz H.D.,

```

```

RA Zachau H.G.;
RT "Subgroup IV of human Immunoglobulin K light chains is encoded by a
RT single germline gene."
RL Nucleic Acids Res. 13:6515-6529(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Z00022; CAA77317.1; -
CC PIR: A01904; KAHUJ1.
CC HSSP: P01789; 1MCP.
CC InterPro: IPR003006; IG_MHC.
CC InterPro: IPR003596; IG_V.
CC Pfam: PF00047; 1g; 1.
CC SMART: SM00406; 1g; 1.
CC Immunoglobulin V region; Signal.
KW SIGNAL
FT CHAIN 1 20 IG KAPPA CHAIN V-IV REGION J1.
FT DOMAIN 21 43 FRAMEWORK-1.
FT DOMAIN 44 60 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 61 75 FRAMEWORK-2.
FT DOMAIN 76 82 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 83 114 FRAMEWORK-3.
FT DOMAIN 115 122 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 43 114 BY SIMILARITY.
FT NON_TER 133 133
SQ SEQUENCE 133 AA; 14632 MW; 5FB3953066744AF4 CRC64;

Query Match 100.0%; Score 46; DB 1; Length 133;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLLIYMA 8
Db 70 PKLLIYMA 77

RESULT 4
KV4C_HUMAN STANDARD; PRT; 134 AA.
AC P06314;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-IV region B17 precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=86041854; PubMed=2997713;
RA Marsh P., Mills F., Gould H.;
RT "Detection of a unique human V kappa IV germline gene by a cloned
RT cDNA probe."
RL Nucleic Acids Res. 13:6531-6544(1985).
RN [2]
RP REVISION TO 76.
RA Marsh P.;
RL Submitted (OCT-1986) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: X02990; CAA26733.1; -.
DR PIR: A01905; K4H017.
DR HSSP: P01789; 1MCP.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003596; IG_v.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00406; IgV_1.
DR Immunoglobulin V region; Signal.
KM SIGNAL 1 20
FT CHAIN 1 134 IG KAPPA CHAIN V-IV REGION B17.
FT DOMAIN 21 43 FRAMEWORK-1.
FT DOMAIN 44 60 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 61 75 FRAMEWORK-2.
FT DOMAIN 76 82 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 83 114 FRAMEWORK-3.
FT DOMAIN 115 121 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 122 133 FRAMEWORK-4.
FT DISULFID 43 114 BY SIMILARITY.
FT NON_TER 134 134
SQ SEQUENCE 134 AA; 14966 MW; 6413A22FD073832 CRC64;

Query Match 100.0%; Score 46; DB 1; Length 134;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PKLITVWA 8
| | | | |
Db 70 PKLITVWA 77

RESULT 5
ID YCKC_BACSU STANDARD; PRT; 151 AA.
AC P42401;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 17.0 kDa protein in COM5 region (ORF3).
GN YCKC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124189; PubMed=8969502;
RA Yamane K., Kumano M., Kurita K.;
RT "The 25 degrees-36 degrees region of the Bacillus subtilis
chromosome: determination of the sequence of a 146 kb segment and
RT identification of 113 genes."
RL Microbiology 142:3047-3056(1996).
RN [2]
RP SEQUENCE OF 43-151 FROM N.A.
RC STRAIN=168;
RX MEDLINE=95219080; PubMed=7704255;
RA Fujiishima Y., Yamane K.;
RT "A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)
of stfAA of the Bacillus subtilis chromosome."
RL Microbiology 141:277-279(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
CC -----
DR EMBL: D50453; BAA08973.1; -.
DR EMBL: D30762; BAA06427.1; -.
DR EMBL: Z59105; CAB12133.1; -.
DR Subtilist; BG11179; yckC.
KM Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 45 65 POTENTIAL.
FT TRANSMEM 91 111 POTENTIAL.
SQ SEQUENCE 151 AA; 16983 MW; 55395791ACD3B0D CRC64;

Query Match 76.1%; Score 35; DB 1; Length 151;
Best Local Similarity 85.7%; Pred. No. 5.5;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKLITVW 7
| | | | |
Db 26 PKLITVW 32

RESULT 6
ID YOR5_ADEG1 STANDARD; PRT; 283 AA.
AC P20747;
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 31.5 kDa protein (ORF 5) (ORF10).
OS Avian adenovirus gall (strain Phelps) (Fowl adenovirus 1) (CELO).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Avadenovirus.
OX NCBI_TaxID=10553;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90251474; PubMed=2160072;
RA Akopian T.A., Kruglyak V.A., Rivkina M.B., Naroditsky B.S.,
RA Tikhonenko T.I.;
RT "Sequence of an avian adenovirus (CELO) DNA fragment (0-11.2%).";
RL Nucleic Acids Res. 18:2825-2825(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96186720; PubMed=8627769;
RA Chlocea S., Kutzbauer R., Schaffner G., Baker A., Mautner V.,
RA Cotten M.;
RT "The complete DNA sequence and genomic organization of the avian
adenovirus CELO."
RL J. Virol. 70:2939-2949(1996).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: X17217; CAA35087.1; -.
DR EMBL: U46933; AAC54931.1; -.
DR PIR: S10005; S10005.
KM Hypothetical protein
SQ SEQUENCE 283 AA; 31487 MW; 6018412DA598183D CRC64;

Query Match 76.1%; Score 35; DB 1; Length 283;
Best Local Similarity 75.0%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PKLITVWA 8
| | | | |
Db 169 PKLITVWA 176

RESULT 7
```

```

PURL_PYRHO
ID PURL_PYRHO STANDARD: PRT: 705 AA.
AC 059621;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3) (FGAM
DE synthase II).
DE PURL OR PH1953.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarayashi Y., Sawada M., Horikawa H., Hatakeyama A., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Oguni A.,
RA Funahashi T., Tanaka T., Kudo Y., Imaizaki Y., Kusuda N., Ouchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
CC -1- CATALYTIC ACTIVITY: ATP + 5'-phosphoribosylformylglycinamide + L-
CC glutamine + H2O = ADP + phosphate + 5'-phosphoribosylformylglycinamide.
CC -1- PATHWAY: DE NOVO PURINE BIOSYNTHESIS; FOURTH STEP.
CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS, PURO AND PURL.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PFAMS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AP000007; BAA31080.1; .
DR InterPro: IPR000728; AIRS-related.
DR Pfam: PF00586; AIRS_C.2.
DR Pfam: PF02769; AIRS_C.2.
KW Purine biosynthesis; Ligase; ATP-binding; Complete proteome.
FT NP_BIND 89 100
FT MOD_RES 100
SQ SEQUENCE 705 AA; 78540 MW; 5750733AAA4DE29 CRC64;

Query Match
Best local Similarity 76.1%; Score 35; DB 1; Length 705;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKLLITW 7
   11 : 11
Db 334 PKPIVW 340

RESULT 8
AP4A_HUMAN
ID AP4A_HUMAN STANDARD: PRT: 146 AA.
AC P50583;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bis(5'-nucleosyl)-tetraphosphatase (Asymmetrical) (EC 3.6.1.17)
DE (Diadenosine 5',5'-P1,P4-tetraphosphate asymmetrical hydrolase)
DE (Diadenosine tetraphosphatase) (AP4A hydrolase) (AP4AASE).
DE NUDIX OR APAH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver, and Spleen;
RX MEDLINE=96067583; PubMed=7487923;
RA Thorne N.M.H., Hankin S., Wilkinson M.C., Nunez C.,
RA Barricough R., McLennan A.G.;
RT "Human diadenosine 5',5'-P1,P4-tetraphosphate pyrophosphohydrolase
RT is a member of the MGT family of nucleotide pyrophosphatases."
RL Biochem. J. 311:717-721(1995).
RN [2]
RP ACETYLATION.
RA McLennan A.G.;
RL Submitted (JAN-1997) to the SWISS-PROT data bank.
CC -1- FUNCTION: ASYMMETRICALLY HYDROLYZES AP4A TO YIELD AMP AND ATP.
CC -1- CATALYTIC ACTIVITY: P(1),P(4)-bis(5'-adenosyl)tetraphosphate +
CC H2O = ATP + AMP.
CC -1- SIMILARITY: BELONGS TO THE NUDIX HYDROLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U30313; AAC50277.1; .
DR MIM: 602852; .
DR InterPro: IPR000086; NUDIX_hydrolase.
DR Pfam: PF00293; NUDIX_1.
DR PRINTS: PR00502; NUDIXFAMILY.
DR PROSITE: PS00893; NUDIX_1.
KW Hydrolase; Acetylation.
FT INIT_MET 0
FT DOMAIN 42 63 NUDIX BOX.
FT MOD_RES 1 1 ACETYLATION.
FT MOD_RES 1 1
SQ SEQUENCE 146 AA; 16698 MW; 7C4D1B42CA87AB8 CRC64;

Query Match
Best local Similarity 73.9%; Score 34; DB 1; Length 146;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKLLITW 7
   11 : 11
Db 87 PKPIVW 93

RESULT 9
AP4A_MOUSE
ID AP4A_MOUSE STANDARD: PRT: 146 AA.
AC P56380;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bis(5'-nucleosyl)-tetraphosphatase (Asymmetrical) (EC 3.6.1.17)
DE (Diadenosine 5',5'-P1,P4-tetraphosphate asymmetrical hydrolase)
DE (Diadenosine tetraphosphatase) (AP4A hydrolase) (AP4AASE).
DE NUDIX OR APAH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,
RA Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,
RA Theisinger B., Wylie T., Lennon G., Soares B., Wilson R., Waterson R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ASYMMETRICALLY HYDROLYZES AP4A TO YIELD AMP AND ATP (BY
CC SIMILARITY).

```



```

CC -1- CATALYTIC ACTIVITY: P(1),P(4)-bis(5'-adenosyl)tetrphosphate +
CC H(2'O - ATP + AMP.
CC -1- SIMILARITY: BELONGS TO THE NUDIX HYDROLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AA125383; -; NOT_ANNOTATED_CDS.
DR EMBL; AA222229; -; NOT_ANNOTATED_CDS.
DR EMBL; AA222233; -; NOT_ANNOTATED_CDS.
DR InterPro: IPR000086; NUDIX_hydrolase.
DR Pfam: PF00293; NUDIX.1
DR PRINTS; PR00502; NUDIXFAMILY.
DR PROSITE; PS00893; NUDIX.1.
KW Hydrolase; Acetylation.
FT INIT_MET 0 BY SIMILARITY.
FT DOMAIN 42 63 NUDIX BOX.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
SQ SEQUENCE 146 AA; 16943 MW; 333AEFAD2E76671 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 146;
Best Local Similarity 71.4%; Pred. No. 8.3;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKLLIYW 7
Db 87 PKVIYW 93

RESULT 10
AP4A_PIG STANDARD; PRT; 146 AA.
ID AP4A_PIG
AC P50584;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bis(5'-nucleosyl)-tetrphosphate (Asymmetrical) (EC 3.6.1.17)
DE (Diadenosine 5',5'-P1,P4-tetrphosphate asymmetrical hydrolase)
DE (Diadenosine tetrphosphate) (AP4A hydrolase) (AP4Aase).
GN NUDT2 OR APAH1.
OS Eukaryota (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Small Intestine;
RA Hankin S., Winteroe A.K., McLennan A.G.;
RA Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: ASYMMETRICALLY HYDROLYSES APA4 TO YIELD AMP AND ATP.
CC -1- CATALYTIC ACTIVITY: P(1),P(4)-bis(5'-adenosyl)tetrphosphate +
CC H(2'O = ATP + AMP.
CC -1- SIMILARITY: BELONGS TO THE NUDIX HYDROLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U38619; AAB61380.1; -
DR InterPro: IPR000086; NUDIX_hydrolase.
DR Pfam; PF00293; NUDIX.1.
DR PRINTS; PR00502; NUDIXFAMILY.
DR PROSITE; PS00893; NUDIX.1.

```

```

KW Hydrolase; Acetylation.
FT INIT_MET 0 BY SIMILARITY.
FT DOMAIN 42 63 NUDIX BOX.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
SQ SEQUENCE 146 AA; 16704 MW; 7AFEC559CBEBA379 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 146;
Best Local Similarity 71.4%; Pred. No. 8.3;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKLLIYW 7
Db 87 PKVIYW 93

RESULT 11
RBP2_HUMAN STANDARD; PRT; 3224 AA.
ID RBP2_HUMAN
AC P49792; Q15280;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Ran-binding protein 2 (RanBP2) (Nuclear pore complex protein Nup358)
DE (Nucleoporin Nup358) (358 kDa nucleoporin) (P270).
GN RANBP2 OR NUP358.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95294031; Pubmed=7775481.
RA Wu J., Martinis M.J., Kraemer D., Blobel G., Couvras E.;
RA "Nup358, a cytoplasmically exposed nucleoporin with peptide repeats,
RA Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous
RA domain, and a leucine-rich region."
RA J. Biol. Chem. 270:14209-14213(1995).
RN [2]
RP SEQUENCE FROM N.A.
RP TISSUE=Blood;
RX MEDLINE=95327194; Pubmed=7603572;
RA Yokoyama N., Hayashi N., Seki T., Nishii K., Hayashida T.,
RA Kuma K.T., Miyata T., Fukui M., Nishimoto T., Pante N., Aebi U.;
RA "A giant nucleoporin protein that binds Ran/TC4."
RA Nature 376:184-188(1995).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.96 ANGSTROMS) OF 1171-1304.
RX MEDLINE=99176415; Pubmed=10078529;
RA Vetter I.R., Nowak G., Nishimoto T., Kuhlmann J., Wittinghofer A.;
RA "Structure of a Ran-binding domain complexed with Ran bound to a GTP
RA analogue: implications for nuclear transport."
RA Nature 398:39-46(1999).
CC -1- FUNCTION: INVOLVED IN TRANSPORT FACTOR (RAN-GTP, KARYOPHERIN) -
CC MEDIATED PROTEIN IMPORT VIA THE F-G REPEAT-CONTAINING DOMAIN WHICH
CC ACTS AS A DOCKING SITE FOR SUBSTRATES. COULD ALSO HAVE ISOMERASE
CC OR CHAPERONE ACTIVITY AND MAY BIND RNA OR DNA. COMPONENT OF THE
CC NUCLEAR EXPORT PATHWAY. SPECIFIC DOCKING SITE FOR THE NUCLEAR
CC EXPORT FACTOR EXPORTIN-1.
CC -1- SUBUNIT: FORMS A TIGHT COMPLEX IN ASSOCIATION WITH RANBP1 AND THE
CC UBICUITIN-CONUGATING ENZYME E2 (UBC9) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX, CYTOPLASMIC FILAMENTS.
CC -1- DOMAIN: CONTAINS MANY X-X-F-G AND X-F-X-F-G REPEATS.
CC -1- SIMILARITY: CONTAINS 4 RANBP1-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 8 RANBP2-TYPE ZINC FINGERS.
CC -1- SIMILARITY: CONTAINS 1 CYCLOPHILIN-LIKE PIRASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```



OY 1 PKLITYWA 8  
| | | | |  
DB 44 PKLITYGA 51

RESULT 14  
KV5R MOUSE STANDARD: PRT: 108 AA.  
ID P01651:  
DT 21-JUL-1986 (rel. 01, Created)  
DT 21-JUL-1986 (rel. 01, Last sequence update)  
DT 15-JUL-1999 (rel. 38, Last annotation update)  
DE Ig kappa chain V-V region EPC 109.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE:  
RX MEDLINE=79195288; PubMed=109517;  
RA Vrana M., Rudikoff S., Potter M.;  
RT "The structural basis of a hapten-inhibitable kappa-chain idiotype."  
RL J. Immunol. 122:1905-1910(1979).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT  
CC BIND BETA(2-1)-FRUCTOFURANOSYL MOIETIES (INULIN).  
CC PIR: B92808; KYMS09.  
DR HSSP: P01607; IRET.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003596; Ig\_V.  
DR Pfam: PF00047; Ig; 1.  
DR SMART: SM00406; IGV; 1.  
KW Immunoglobulin V region.  
FT DOMAIN 1 23 FRAMEWORK-1.  
FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.  
FT DOMAIN 35 49 FRAMEWORK-2.  
FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.  
FT DOMAIN 57 88 FRAMEWORK-3.  
FT DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-3.  
FT DOMAIN 98 108 FRAMEWORK-4.  
FT DISULFID 23 88 BY SIMILARITY.  
FT NON\_TER 108 108  
SQ SEQUENCE 108 AA; 11876 MW; 35C116BD0F79310 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 108;  
Best Local Similarity 87.5%; Pred. No. 9.6;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKLITYWA 8  
| | | | |  
DB 44 PKLITYGA 51

RESULT 15  
KV5S MOUSE STANDARD: PRT: 108 AA.  
ID P01652:  
DT 21-JUL-1986 (rel. 01, Created)  
DT 21-JUL-1986 (rel. 01, Last sequence update)  
DT 15-JUL-1999 (rel. 38, Last annotation update)  
DE Ig kappa chain V-V region J606.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE:  
RX MEDLINE=82099361; PubMed=6798111;  
RA Johnson N., Stankard J., Paul L.;  
RT "The complete V domain amino acid sequences of two myeloma inulin-  
binding proteins."  
RL J. Immunol. 128:302-307(1982).

CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT  
CC BIND BETA(2-1)-FRUCTOFURANOSYL MOIETIES (INULIN).  
CC PIR: A92811; KYMS06.  
DR HSSP: P01607; IRET.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003596; Ig\_V.  
DR Pfam: PF00047; Ig; 1.  
DR SMART: SM00406; IGV; 1.  
KW Immunoglobulin V region.  
FT DOMAIN 1 23 FRAMEWORK-1.  
FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.  
FT DOMAIN 35 49 FRAMEWORK-2.  
FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.  
FT DOMAIN 57 88 FRAMEWORK-3.  
FT DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-3.  
FT DOMAIN 98 108 FRAMEWORK-4.  
FT DISULFID 23 88 BY SIMILARITY.  
FT NON\_TER 108 108  
SQ SEQUENCE 108 AA; 11810 MW; 8DE4DD31076F2AFB CRC64;

Query Match 71.7%; Score 33; DB 1; Length 108;  
Best Local Similarity 87.5%; Pred. No. 9.6;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKLITYWA 8  
| | | | |  
DB 44 PKLITYGA 51

Search completed: July 15, 2002, 13:23:32  
Job time: 1448 sec

✓

,

,

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:22:42 ; Search time 172.49 Seconds  
(Without alignments)  
8.023 Million cell updates/sec

Title: US-09-712-819A-4  
Perfect score: 46  
Sequence: 1 PKLIIYMA 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_19:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.todent:\*  
12: sp.virus:\*  
13: sp.vertelbrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	46	100.0	130	4	Q9NP29	Q9NP29 homo sapien
2	41	89.1	2703	10	Q9MAZ4	Q9MAZ4 arabidopsis
3	41	89.1	2810	10	Q9FKS4	Q9FKS4 arabidopsis
4	37	80.4	99	11	Q9UL74	Q9UL74 mus musculu
5	37	80.4	112	9	Q38357	Q38357 lactococcus
6	37	80.4	227	10	Q23209	Q23209 arabidopsis
7	36	80.4	375	16	Q92KS3	Q92KS3 rhizobium m
8	36	78.3	291	11	Q35558	Q35558 mus musculu
9	35	76.1	288	5	Q9VRU6	Q9VRU6 drosophila
10	35	76.1	329	6	Q97885	Q97885 equus cabal
11	35	76.1	362	17	Q9V216	Q9V216 pyrococcus
12	34	73.9	147	11	Q9B6V2	Q9B6V2 mus musculu
13	34	73.9	147	11	Q9D2U6	Q9D2U6 mus musculu
14	34	73.9	202	12	Q9IBM3	Q9IBM3 avian adeno
15	34	73.9	303	12	Q64777	Q64777 avian adeno
16	34	73.9	367	10	Q9STR5	Q9STR5 arabidopsis

17	34	73.9	401	10	Q64859	Q64859 arabidopsis
18	34	73.9	438	2	Q944H9	Q944H9 arabidopsis
19	34	73.9	402	2	Q87856	Q87856 streptomyce
20	34	73.9	496	17	Q59477	Q59477 pyrococcus
21	34	73.9	556	12	Q36369	Q36369 alcelaphine
22	34	73.9	639	17	Q59084	Q59084 pyrococcus
23	34	73.9	706	2	Q88090	Q88090 enterococcu
24	34	73.9	815	16	Q9KLG3	Q9KLG3 vibrio chol
25	34	73.9	904	4	Q9H0B2	Q9H0B2 homo sapien
26	34	73.9	1666	5	Q917P4	Q917P4 drosophila
27	34	73.9	1765	4	Q99666	Q99666 homo sapien
28	34	73.9	3053	11	Q9ER09	Q9ER09 mus musculu
29	33.5	72.8	438	10	Q94A82	Q94A82 arabidopsis
30	33	71.7	111	11	Q920E9	Q920E9 mus musculu
31	33	71.7	244	10	Q9LQW2	Q9LQW2 arabidopsis
32	33	71.7	256	16	Q67471	Q67471 equifex aeo
33	33	71.7	519	16	Q9KTN8	Q9KTN8 vibrio chol
34	33	71.7	527	16	Q9K1O9	Q9K1O9 vibrio chol
35	33	71.7	687	3	Q74520	Q74520 schizosacch
36	33	71.7	817	5	Q9VZJ5	Q9VZJ5 drosophila
37	33	71.7	939	16	Q9PLB7	Q9PLB7 chlamydia m
38	33	71.7	956	16	Q84812	Q84812 chlamydia t
39	33	71.7	1483	3	Q96UL6	Q96UL6 diaporthe a
40	32	69.6	103	11	Q9JL80	Q9JL80 mus musculu
41	32	69.6	107	4	Q96SA9	Q96SA9 homo sapien
42	32	69.6	110	16	Q84455	Q84455 chlamydia t
43	32	69.6	153	10	Q04520	Q04520 arabidopsis
44	32	69.6	183	4	Q9Y2V0	Q9Y2V0 homo sapien
45	32	69.6	217	17	Q27742	Q27742 methanochet

## ALIGNMENTS

RESULT 1  
Q9NP29 1 PRELIMINARY; PRT; 130 AA.  
ID Q9NP29  
AC Q9NP29;  
DP 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DR 01-DEC-2000 (TREMBLrel. 19, Last annotation update)  
DE MICROFILBRILLAR PROTEIN 2 (FRAGMENT).  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96354815; PubMed=8753791;  
RA Ozsvath K.J., Xia S., Hirose H., Tilson M.D.;  
RT "Two hypothetical proteins of human aortic adventitia, with Ig kappe, collagenous, and aromatic-rich motifs."  
RT Biochem. Biophys. Res. Commun. 225:500-504(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97367690; PubMed=9224393;  
RA Ozsvath K.J., Hirose H., Xia S., Chew D., Knoetgen J. III,  
RA Tilson M.D.;  
RT "Expression of two novel recombinant proteins from aortic adventitia (kappatids) sharing amino acid sequences with cytomegalovirus."  
RT J. Surg. Res. 69:277-282(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Ozsvath K.J., Xia S., Hirose H., Tilson M.D.;  
RA Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF206020; AAF62402.1; -  
DR HSSP; P80362; IWTL.  
DR InterPro; IPR003600; Iq\_1ike.  
DR SMART; SM00410; IG\_1like; 1.  
FT NON-TER 130  
SQ SEQUENCE 130 AA; 14128 MW; 51275185AC6FAIE CRC64;

```

Query Match      100.0%; Score 46; DB 4; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLLIYW 8
Db 70 PKLLIYW 77

RESULT 2
O9MAZ4 PRELIMINARY; PRT; 2703 AA.
ID O9MAZ4
AC O9MAZ4
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ARAD3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=COLUMBIA;
RC Sugiyama H., Oguchi K., Tamura K., Takahashi H.;
RT "Arabidopsis thaliana mRNA for ARAD3, complete cds.";
RL EMBL: AB040133; BAA92828.1; -.
DR InterPro: IPR003151; FATC.
DR InterPro: IPR003152; FATC.
DR InterPro: IPR000403; P13_P14_Kinase.
DR Pfam: PF02259; FAT; 1.
DR Pfam: PF02260; FATC; 1.
DR Pfam: PF00454; P13_P14_Kinase; 1.
DR SMART: SM00146; P13KC; 1.
DR PROSITE: PS0290; P13_4_KINASE_3; 1.
SQ SEQUENCE 2703 AA; 302511 MW; B171F2P1EA12AD04 CRC64;

Query Match      89.1%; Score 41; DB 10; Length 2703;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLLIYW 7
Db 972 PKLLIYW 978

RESULT 3
O9FKS4 PRELIMINARY; PRT; 2810 AA.
ID O9FKS4
AC O9FKS4
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ARAD3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=COLUMBIA;
RC MEDLINE=98344145; PubMed=9679202;
RA Kaneo T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence
RT features of the regions of 1,381,565 bp covered by twenty one
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:131-145(1998).

```

```

DR EMBL: AB011477; BAB1344.1; -.
DR InterPro: IPR003151; FAT.
DR InterPro: IPR003152; FATC.
DR InterPro: IPR002016; Peroxidase.
DR InterPro: IPR000403; P13_P14_Kinase.
DR Pfam: PF02259; FAT; 1.
DR Pfam: PF02260; FATC; 1.
DR Pfam: PF00454; P13_P14_Kinase; 1.
DR SMART: SM00146; P13KC; 1.
DR PROSITE: PS00435; PEROXIDASE_1; UNKNOWN_1.
DR PROSITE: PS0290; P13_4_KINASE_3; 1.
SQ SEQUENCE 2810 AA; 314057 MW; FB82AE7CA60BD0E7 CRC64;

Query Match      89.1%; Score 41; DB 10; Length 2810;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLLIYW 7
Db 971 PKLLIYW 977

RESULT 4
O9UL74 PRELIMINARY; PRT; 99 AA.
ID O9UL74
AC O9UL74
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ANTI-MYOSIN IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION
DE (EPACMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BALE/C;
RC MEDLINE=20448942; PubMed=10992488;
RA Malkiel S., Liao L., Cunningham M.W., Diamond B.;
RT "T-cell-dependent antibody response to the dominant epitope of
RT streptococcal polysaccharide, N-acetyl-glucosamine, is cross-reactive
RT with cardiac myosin.";
RL Infect. Immun. 68:5803-5808(2000).
DR EMBL: AF206032; AAF69330.1; -.
DR HSSP: P80362; IWTI.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00406; IGV; 1.
DR NON_TER 1
FT NON_TER 99
FT SEQUENCE 99 AA; 10939 MW; 3B25D0E78453324 CRC64;

Query Match      80.4%; Score 37; DB 11; Length 99;
Best Local Similarity 87.5%; Pred. No. 10;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLLIYW 8
Db 36 PKLLIYW 43

RESULT 5
O3B357 PRELIMINARY; PRT; 112 AA.
ID O3B357
AC O3B357
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL_13.2 KDA PROTEIN.

```

OS Lactococcus delbrueckii bacteriophage LT-H.  
 OC Viruses.  
 NCBI\_TaxID=12348;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93367750; PubMed=8497043;  
 RA Vasala A., Dupont L., Baumann M., Ritzenthaler P., Alatosava T.;  
 RT "Molecular comparison of the structural proteins encoding gene  
 clusters of two related Lactobacillus delbrueckii bacteriophages.";  
 RL J. Virol. 67:3061-3068(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94237431; PubMed=7514146;  
 RA Mikkonen M., Vancisto J., Alatosava T.;  
 RT "Ribosome binding site consensus sequence of Lactobacillus delbrueckii  
 subsp. lactis bacteriophage LT-H.";  
 RL FEMS Microbiol. Lett. 116:315-320(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95129893; PubMed=7828907;  
 RA Mikkonen M., Alatosava T.;  
 RT "Characterization of the genome region encoding structural proteins of  
 Lactobacillus delbrueckii subsp. lactis bacteriophage LT-H.";  
 RL Gene 151:53-59(1994).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96064414; PubMed=8526515;  
 RA Vasala A., Valikkila M., Caldentey J., Alatosava T.;  
 RT "Genetic and biochemical characterization of the Lactobacillus  
 delbrueckii subsp. lactis bacteriophage LT-H lysin.";  
 RL Appl. Environ. Microbiol. 61:4004-4011(1995).  
 RN EMBL: M96254; AAC00555.1;  
 KM Hypothetical protein.  
 SQ SEQUENCE 112 AA; 13203 MW; 81B615035BA7AFEE CRC64;

Query Match 80.4%; Score 37; DB 9; Length 112;  
 Best Local Similarity 75.0%; Pred. No. 11;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KLLIYWA 8  
 DB 22 PTLIYWA 29

RESULT 6  
 O23209 PRELIMINARY; PRT: 227 AA.  
 AC O23209;  
 DT 01-JAN-1998 (TRENBLREL. 05, Created)  
 DT 01-MAY-1999 (TRENBLREL. 10, Last sequence update)  
 DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)  
 DE HYPOTHEICAL 25.2 KDA. PROTEIN.  
 GN C7A10.640 OR AT4G36720.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Terry N., Vos P., Heljnen L., Mewes H.W., Schueller C.,  
 Chaitwalis N.;  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Z99708; CAB16825.1;  
 DR EMBL: AL161589; CAB80338.1;  
 DR InterPro: IPR004345; TB2\_DP1\_HVA22.  
 DR Pfam: PF03134; TB2\_DP1\_HVA22; 1.  
 KM Hypothetical protein.

SQ SEQUENCE 227 AA; 25231 MW; 14C52A04EEFE3E00 CRC64;

Query Match 80.4%; Score 37; DB 10; Length 227;  
 Best Local Similarity 85.7%; Pred. No. 22;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 KLLIYWA 8  
 DB 67 KLLIYWA 73

RESULT 7  
 O92KS3 PRELIMINARY; PRT: 375 AA.  
 AC O92KS3;  
 DT 01-DEC-2001 (TRENBLREL. 19, Created)  
 DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)  
 DE HYPOTHEICAL SIGNAL PEPTIDE PROTEIN SMC00400.  
 GN SMC00400.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OC NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21368234; PubMed=11474104;  
 RA Gallibert F., Finn T.M., Long S.R., Puehler A., Abola P., Ampe F.,  
 RA Barloy-Hubler F., Barnett M.J., Becker A., Bolstad P., Botte G.,  
 RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chai P.,  
 RA Cowie A., Davis R.W., Dreano S., Federici N.A., Fisher R.F.,  
 RA Gloux S., Godie T., Goffeau A., Golding B., Gouzy J., Gurjel M.,  
 RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,  
 RA Kahn M.L., Kaiman S., Keating D.H., Kiss E., Komp C., Lelaure V.,  
 RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetle D., Purnelle B.,  
 RA Rasperger U., Surzycki R., Thebault P., Vandenbol M.,  
 RA Vorhoeper F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;  
 RT "The composite genome of the legume symbiont Sinorhizobium meliloti.";  
 RL Science 293:668-672(2001).  
 DR EMBL: AL591783; CAC41757.1;  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 375 AA; 39924 MW; 79592B7A7A7673B28 CRC64;

Query Match 80.4%; Score 37; DB 16; Length 375;  
 Best Local Similarity 85.7%; Pred. No. 35;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KLLIYWA 7  
 DB 264 KLLIYWA 270

RESULT 8  
 O35558 PRELIMINARY; PRT: 291 AA.  
 AC O35558;  
 DT 01-JAN-1998 (TRENBLREL. 05, Created)  
 DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)  
 DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)  
 DE ERK2.  
 GN MAPK1 OR ERK2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97407954; PubMed=9261178;  
 RX Sugita N., Suga T., Ozeki Y., Mamiya G., Takishima K.;  
 RT "The mouse extracellular signal-regulated kinase 2 gene: gene

RT structure and characterization of the promoter."  
 RL J. Biol. Chem. 272:21575-21581(1997).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL: D872271: BAA22620.1; -  
 DR EMBL: D87264: BAA22620.1; JOINED.  
 DR EMBL: D87265: BAA22620.1; JOINED.  
 DR EMBL: D87266: BAA22620.1; JOINED.  
 DR EMBL: D87267: BAA22620.1; JOINED.  
 DR EMBL: D87268: BAA22620.1; JOINED.  
 DR EMBL: D87269: BAA22620.1; JOINED.  
 DR HSSP: P27703: 1ERK.  
 DR MGD: MGI:1346958: Mapk1.  
 DR InterPro: IPR000719: Euk\_pkinase.  
 DR InterPro: IPR003527: Map\_kin.  
 DR InterPro: IPR002290: Ser\_thr\_pkinase.  
 DR Pfam: PF00069: pkinase; 1.  
 DR SMART: SM00220: S\_TKC; 1.  
 DR PROSITE: PS01351: MAPK; UNKNOWN\_1.  
 DR PROSITE: PS00107: PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00117: PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108: PROTEIN\_KINASE\_ST; 1.  
 KW APP-binding, serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 291 AA; 33601 MW; 7088216f35a8a260 CRC64;

Query Match 78.3%; Score 36; DB 11; Length 291;  
 Best Local Similarity 85.7%; Pred. No. 43;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKLIIYW 7  
 |||  
 Db 282 PKLIITW 288

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacheco J.M.,  
 RA Palazzolo M., Platten G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Stryker R., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AR003566: AAF50779.1; -  
 DR FlyBase: FBgn0035606: CG15214.  
 DR InterPro: IPR003598: 1g\_c2.  
 DR InterPro: IPR003006: 1g\_mhc.  
 DR Pfam: PF00047: 1g; 1.  
 DR SMART: SM00408: IGC2; 1.  
 KW Immunoglobulin domain.  
 SQ SEQUENCE 288 AA; 31235 MW; 8366f44c601617fe CRC64;

Query Match 76.1%; Score 35; DB 5; Length 288;  
 Best Local Similarity 71.4%; Pred. No. 65;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKLIIYW 7  
 |||  
 Db 27 PKLIITW 33

RESULT 9  
 ID 09VRL6 PRELIMINARY; PRT; 288 AA.  
 AC 09VRL6:  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CG15214 PROTEIN.  
 GN CG15214.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Chame M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-D., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos R., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.D., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaisli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RESULT 10  
 ID 097885 PRELIMINARY; PRT; 329 AA.  
 AC 097885:  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE TYROSINASE-RELATED PROTEIN 1 (FRAGMENT).  
 GN TRP1.  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 OX NCBI\_TaxID=9796;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SKIN;  
 RA Rieder S., Checa-Cortes M.L., Joerg H., Stranzinger G.;  
 RT An equine sequence homologous to tyrosinase-related protein-1 (TRP1)  
 RL mapped to chromosome 23/14-16."  
 DR EMBL: AF076781: AAC97109.1; -  
 DR InterPro: IPR002227: Tyrosinase.  
 DR Pfam: PF00264: tyrosinase; 2.  
 DR PRINTS: PR00092: TYROSINASE.  
 DR PROSITE: PS00497: TYROSINASE\_1; 1.  
 FT NON-TER 1  
 FT NON-TER 329  
 SQ SEQUENCE 329 AA; 37447 MW; 1258855FA453P864 CRC64;

Query Match 76.1%; Score 35; DB 6; Length 329;  
 Best Local Similarity 50.0%; Pred. No. 74;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLIITW 8  
 |||  
 Db 212 PKLIITW 219



```

RESULT 11
Q9V2L6 PRELIMINARY; PRT; 362 AA.
ID Q9V2L6;
AC Q9V2L6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DE 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE DOLICHOLOPHOSPHATE MANNOSYLTRANSFERASE.
GN PAB2307.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP STRAIN=ORSAY;
RC
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248283; CAB48982.1;
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Complete proteome.
SQ
SEQUENCE 362 AA; 41505 MW; 50A99F7195F540DE CRC64;

```

```

Query Match 76.1%; Score 35; DB 17; Length 362;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 KLIYWA 8
DB 323 QLIYWA 329

```

```

RESULT 12
Q9D6V2 PRELIMINARY; PRT; 147 AA.
ID Q9D6V2;
AC Q9D6V2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DE 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 2310051106RIK PROTEIN.
GN 2310051106RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP STRAIN=C57BL/6J; TISSUE=TONGUE;
RC MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schirral L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";

```

```

RL Nature 409:685-690(2001).
DR EMBL; AK009933; BAB26592.1;
DR MGD; MGI:1913651; 2310051106RIK.
DR InterPro; IPR000086; NUDIX_hydrolase.
DR PRINTS; PR00502; NUDIXFAMILY.
DR PROSITE; PS00893; NUDIX; 1.
SQ
SEQUENCE 147 AA; 16989 MW; A634D19F1E8CE34 CRC64;

```

```

Query Match 73.9%; Score 34; DB 11; Length 147;
Best Local Similarity 71.4%; Pred. No. 54;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 KLIYWA 7
DB 88 KTIYWA 94

```

```

RESULT 13
Q9D2U6 PRELIMINARY; PRT; 147 AA.
ID Q9D2U6;
AC Q9D2U6;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DE 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 2310051106RIK PROTEIN.
GN 2310051106RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP STRAIN=C57BL/6J; TISSUE=CEREBELLUM;
RC MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schirral L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";

```

```

RL Nature 409:685-690(2001).
DR EMBL; AK018771; BAB31399.1;
DR MGD; MGI:1913651; 2310051106RIK.
DR InterPro; IPR000086; NUDIX_hydrolase.
DR PRINTS; PR00502; NUDIXFAMILY.
DR PROSITE; PS00893; NUDIX; 1.
SQ
SEQUENCE 147 AA; 16957 MW; AFD3C5C71E8CDA64 CRC64;

```

```

Query Match 73.9%; Score 34; DB 11; Length 147;
Best Local Similarity 71.4%; Pred. No. 54;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 KLIYWA 7
DB 88 KTIYWA 94

```

```

RESULT 14

```

```

09IBM3
ID 09IBM3      PRELIMINARY;      PRT:      202 AA.
AC 09IBM3;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ORF18.
OS Avian adenovirus gall (strain Phelps) (Fowl adenovirus 1) (CELO).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.
OX NCBI_TaxID=10553;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PHELPS (ATCC VR-432);
RX MEDLINE=96186720; PubMed=8627769;
RA Chioocca S., Kurzbauner R., Schaffner G., Baker A., Mautner V.,
RA Colten M.;
RT "The complete DNA sequence and genomic organization of the avian
RT adenovirus CELO.";
RL J. Virol. 70:2839-2949(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PHELPS (ATCC VR-432);
RA Colten M.L.;
RL Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL: U46933; AAC54925.1; -
SO SEQUENCE      202 AA: 23210 MW: C0177BD547D2F67 CRC64;

```

```

Query Match      73.9%; Score 34; DB 12; Length 202;
Best Local Similarity 57.1%; Pred. No. 72;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 PKLIYV 7
      |||:::|
Db      39 PKLIYVW 45

```

```

RESULT 15
Q64777      PRELIMINARY;      PRT:      303 AA.
ID 064777;
AC 064777;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ORF5 PROTEIN (FRAGMENT)
OS Avian adenovirus gall (strain Phelps) (Fowl adenovirus 1) (CELO).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.
OX NCBI_TaxID=10553;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CELO;
RA Akopian T.A., Kaverina E.N., Kruglyak V.A., Naroditsky B.S.,
RA Tikhonenko T.T.;
RT "Sequence of an Avian Adenovirus (CELO) DNA Fragment (11.2 - 19.2
RT kb).";
RL Submitted (MAY-1993) to the EMBL/Genbank/DBJ databases.
DR EMBL: Z22864; CAA80481.1; -
FT NON_TER      303
FT SEQUENCE      303 AA: 34492 MW: 0A34B80F941D513C CRC64;
SO

```

```

Query Match      73.9%; Score 34; DB 12; Length 303;
Best Local Similarity 57.1%; Pred. No. 1.1e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 PKLIYV 7
      |||:::|
Db      39 PKLIYVW 45

```

Search completed: July 15, 2002, 13:22:44  
 Job time: 1489 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: July 15, 2002, 12:57:56 ; Search time 228.39 seconds  
(without alignments)  
3.891 Million cell updates/sec

Title: US-09-712-819A-4  
Perfect score: 46  
Sequence: 1 PKLITWA 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

## Database :

A.Geneseq.032802:\*

1:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1980.DAT:*
2:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1981.DAT:*
3:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1982.DAT:*
4:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1983.DAT:*
5:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1984.DAT:*
6:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1985.DAT:*
7:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1986.DAT:*
8:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1987.DAT:*
9:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1988.DAT:*
10:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1989.DAT:*
11:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1990.DAT:*
12:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1991.DAT:*
13:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1992.DAT:*
14:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1993.DAT:*
15:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1994.DAT:*
16:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1995.DAT:*
17:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1996.DAT:*
18:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1997.DAT:*
19:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1998.DAT:*
20:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA1999.DAT:*
21:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA2000.DAT:*
22:	/SIDSI/gcgdata/hold-geneseq/genesqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	100.0	82	19	AAW62807
2	46	100.0	84	17	AAW14491
3	46	100.0	84	17	AAW9878
4	46	100.0	94	19	AAW62808
5	46	100.0	100	19	AAW70948
6	46	100.0	106	19	AAW31647
7	46	100.0	107	13	AAW25723
8	46	100.0	107	14	AAW38599
9	46	100.0	107	19	AAW80817
10	46	100.0	107	19	AAW80814
11	46	100.0	107	19	AAW58499

12	46	100.0	107	19	AAW47087
13	46	100.0	107	20	AAW17995
14	46	100.0	107	21	AAW90375
15	46	100.0	107	21	AAW79425
16	46	100.0	107	22	AAW69665
17	46	100.0	107	22	AAW69666
18	46	100.0	107	22	AAW69666
19	46	100.0	110	13	AAW30012
20	46	100.0	110	19	AAW65520
21	46	100.0	110	19	AAW65517
22	46	100.0	111	15	AAW52059
23	46	100.0	112	13	AAW28807
24	46	100.0	112	15	AAW4932
25	46	100.0	112	15	AAW54933
26	46	100.0	113	15	AAW50313
27	46	100.0	113	15	AAW50314
28	46	100.0	113	15	AAW50322
29	46	100.0	113	15	AAW59513
30	46	100.0	113	15	AAW56964
31	46	100.0	113	15	AAW53770
32	46	100.0	113	15	AAW63769
33	46	100.0	113	17	AAW92215
34	46	100.0	113	17	AAW92217
35	46	100.0	113	18	AAW27695
36	46	100.0	113	20	AAW50145
37	46	100.0	113	20	AAW42267
38	46	100.0	113	20	AAW42268
39	46	100.0	113	20	AAW42269
40	46	100.0	113	20	AAW57177
41	46	100.0	113	21	AAW18855
42	46	100.0	113	21	AAW18857
43	46	100.0	113	21	AAW18861
44	46	100.0	113	21	AAW18863
45	46	100.0	113	21	AAW18863

## ALIGNMENTS

RESULT	ID	AAW62807	standard; Peptide; 82 AA.
XX	XX	AAW62807	
XX	XX	23-SEP-1998 (first entry)	
XX	XX	Amino acid sequence of a human antibody fragment.	
XX	XX	Human; immunoglobulin; Ig; transgenic; non-human mammal;	
XX	XX	Inactivated endogenous Ig locus; B-cell development;	
XX	XX	human heavy chain Ig locus; micro constant region; J-H; D-H; V-H gene;	
XX	XX	kappa light chain Ig locus; kappa constant region; J-kappa gene; V-kappa;	
XX	XX	production; antibody.	
XX	XX	Homo sapiens.	
XX	XX	OS	
XX	XX	W09824893-A2.	
XX	XX	11-JUN-1998.	
XX	XX	03-DEC-1997; 97WO-US23091.	
XX	XX	03-DEC-1996; 96US-0759620.	
XX	XX	(ABGE-) ABGENIX INC.	
XX	XX	Green J., Jakobovits A., Klapholz S., Kucherlapati R;	
XX	XX	Mendez M;	
XX	XX	WPI: 1998-333314/29.	
XX	XX	New transgenic non-human mammals - having an inactivated	

immunoglobulin locus and a near complete human immunoglobulin locus,  
 used for production of human antibodies  
 Disclosure; Page 78; 128pp; English.  
 AAW62793-822 represent fragments of human antibodies produced by  
 transgenic xenomice, created using the method of the invention. The  
 specification describes a transgenic non-human mammal which has genome  
 modifications that comprise an inactivated endogenous immunoglobulin (Ig)  
 locus, so that the mammal does not display normal B-cell development. The  
 modified genome also has an inserted human heavy chain Ig locus in  
 germline configuration, the human heavy chain Ig locus comprising a human  
 micro constant region and regulatory and switch sequences, human J-H  
 genes, human D-H genes, and human V-H genes and an inserted human kappa  
 light chain Ig locus in germline configuration, the human kappa light  
 chain Ig locus comprising a human kappa constant region, J-kappa genes,  
 and V-kappa genes, where the number of V-H and V-kappa genes inserted  
 are selected to restore normal B-cell development in the mammal. The  
 transgenic animals have a near complete human Ig locus, including both a  
 human heavy chain locus and a human kappa light chain locus. They can  
 be used for the production of human antibodies when exposed to  
 particular antigens e.g. when exposed to human IL-8, EGFR or TNF- alpha  
 the mice will produce antibodies to IL-8, EGFR or TNF- alpha  
 respectively.  
 Sequence 82 AA;  
 SQ  
 Query Match 100.0%; Score 46; DB 19; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 0.21;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PKLLIYWA 8  
 |||||  
 Db 31 PKLLIYWA 38  
 RESULT 2  
 AAW14491  
 ID AAW14491 standard; Protein; 84 AA.  
 AC AAW14491;  
 XX  
 DT 28-JAN-1997 (first entry)  
 XX  
 DE Monoclonal antibody D VK.  
 XX  
 KW heavy; light chain: monoclonal antibody; antigen 3; marker; melanoma;  
 KM permanent human tumour cell line; tumour-associated antigen; epitope;  
 KW gastrointestinal tumour; pancreatic carcinoma; diagnostic; therapeutic;  
 KM antigen 11; Vibrio cholera; neuraminidase-resistant; ganglioside GD2.  
 XX  
 OS Synthetic.  
 XX  
 PN EP727436-A1.  
 XX  
 PD 21-AUG-1996.  
 XX  
 PF 21-MAR-1990; 90EP-0105322.  
 XX  
 PR 24-MAR-1989; 89DE-3909799.  
 XX  
 PA (BEHW ) BEHRINGWERKE AG.  
 XX  
 PI Auerbach B, Bosslet K, Sedlacek H, Seemann G;  
 XX  
 DR WPI: 1996-372835/38.  
 DR N-PSDB; AAT63508.  
 XX  
 PT Monoclonal antibody to tumour-associated antigen - useful as  
 PT gastrointestinal tumour marker  
 XX  
 PS Disclosure; Page 14; 19pp; German.

XX  
 AAW14490-91 are the heavy and light chains (respectively) of monoclonal  
 antibody (Mab) D. Mab D recognises Vibrio cholera  
 CC neuraminidase-resistant epitope of ganglioside GD2, from a human melanoma  
 CC cell line. Mabs A, B and C (see AAW14484-89) are mentioned in the  
 CC specification, but are not part of the claims. Mabs A and B recognise  
 CC antigens 3 and 11 resp., of a permanent human tumour cell line. Mab C  
 CC also recognises an epitope of a tumour-associated antigen. These antigens  
 CC occur at high concns. in the serum of patients with gastrointestinal  
 CC tumours, e.g. pancreatic carcinoma, and are thus useful as tumour markers  
 CC for diagnostic or therapeutic purposes.  
 XX  
 SQ Sequence 84 AA;  
 QY 1 PKLLIYWA 8  
 |||||  
 Db 22 PKLLIYWA 29  
 RESULT 3  
 AAR9878  
 ID AAR9878 standard; Protein; 84 AA.  
 AC AAR9878;  
 XX  
 DT 28-JAN-1997 (first entry)  
 XX  
 DE Monoclonal antibody D VK.  
 XX  
 KW Monoclonal antibody; Mab; epitope; tumour-associated antigen;  
 KW marker; antigen.  
 XX  
 OS Synthetic.  
 XX  
 PN EP727435-A1.  
 XX  
 PD 21-AUG-1996.  
 XX  
 PF 21-MAR-1990; 90EP-0105322.  
 XX  
 PR 24-MAR-1989; 89DE-3909799.  
 XX  
 PA (BEHW ) BEHRINGWERKE AG.  
 XX  
 PI Auerbach B, Bosslet K, Sedlacek H, Seemann G;  
 XX  
 DR WPI: 1996-372835/38.  
 DR N-PSDB; AAT36666.  
 XX  
 PT Monoclonal antibody to tumour-associated antigen - useful as  
 PT gastrointestinal tumour marker  
 XX  
 PS Disclosure; Page 14; 19pp; German.  
 CC Mab C (AAT36659-T36660) is a monoclonal antibody that recognises an  
 CC epitope of a tumour-associated antigen occurring at high concn. in  
 CC the serum of patients with gastrointestinal tumours, e.g. pancreatic  
 CC carcinoma, and is thus useful as a tumour marker for diagnostic or  
 CC therapeutic purposes.  
 CC Mabs A, B and D are mentioned in the specification, but are not  
 CC part of the claims.  
 CC Mab A (AAT36661-T36662) recognises antigen 3 of permanent human  
 CC tumour cell line.  
 CC Mab B (AAT36663-T36664) recognises antigen 11 of permanent human  
 CC tumour cell line.  
 CC Mab D (AAT36665-T36666) recognises a Vibrio cholera neuraminidase-  
 CC resistant epitope of ganglioside GD2, from a human melanoma cell  
 CC line.

XX Sequence 84 AA;  
SQ

Query Match 100.0%; Score 46; DB 17; Length 84;  
Best Local Similarity 100.0%; Pred. No. 0.21;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLLIYWA 8  
|||||||  
Db 22 PKLLIYWA 29

RESULT 4  
AAW62808  
ID AAW62808 standard; Peptide: 94 AA.

XX AAW62808;

XX 23-SEP-1998 (first entry)

XX Amino acid sequence of a human antibody fragment.

XX Human; immunoglobulin; Ig; transgenic; non-human mammal;

XX Inactivated endogenous Ig locus; B-cell development;

XX human heavy chain Ig locus; micro constant region; J-H: D-H: V-H gene;

XX kappa light chain Ig locus; kappa constant region; J-kappa gene; V-kappa;

XX production; antibody.

XX Homo sapiens.

XX WO9824893-A2.

XX 11-JUN-1998.

XX 03-DEC-1997; 97WO-US23091.

XX 03-DEC-1996; 96US-0759620.

XX (ABGE-) ABGENIX INC.

XX Green L, Yakobovits A, Klapholz S, Kucherlapati R;

XX Mendez M;

XX WPI: 1998-333314/29.

XX New transgenic non-human mammals - having an inactivated  
PT immunoglobulin locus and a near complete human immunoglobulin locus,  
PT used for production of human antibodies

XX Disclosure; Page 79; 128pp; English.

XX AAW62793-822 represent fragments of human antibodies produced by

XX transgenic Xenomice, created using the method of the invention. The

XX specification describes a transgenic non-human mammal which has genome

XX modifications that comprise an inactivated endogenous immunoglobulin (Ig)

XX locus, so that the mammal does not display normal B-cell development. The

XX modified genome also has an inserted human heavy chain Ig locus in

XX germline configuration, the human heavy chain Ig locus comprising a human

XX micro constant region and switch sequences, human J-H

XX genes, human D-H genes, and human V-H genes and an inserted human kappa

XX light chain Ig locus in germline configuration, the human kappa light

XX chain Ig locus comprising a human kappa constant region, J-kappa genes,

XX and V-kappa genes, where the number of V-H and V-kappa genes inserted

XX are selected to restore normal B-cell development in the mammal. The

XX transgenic animals have a near complete human Ig locus, including both a

XX human heavy chain locus and a human kappa light chain locus. They can

XX be used for the production of human antibodies when exposed to

XX particular antigens e.g. when exposed to human IL-8, Bcfr or TNF- alpha

XX the mice will produce antibodies to IL-8, Bcfr or TNF- alpha

XX Sequence 94 AA;

Query Match 100.0%; Score 46; DB 19; Length 94;

Best Local Similarity 100.0%; Pred. No. 0.24;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLLIYWA 8  
|||||||  
Db 31 PKLLIYWA 38

RESULT 5  
AAW70948  
ID AAW70948 standard; protein: 100 AA.

XX AAW70948;

XX 14-OCT-1998 (first entry)

XX Light chain of monoclonal antibody RS-348.

XX Pathogenic virus; tropism; mucosa; CDR region; monoclonal antibody;

XX respiratory syncytial virus; RSV; VP6 protein; rota virus; RV;

XX viral infection; inhibit; fusion; protection; transcription;

XX antiviral agent; prophylaxis; diagnosis; infection; contamination.

XX Mus sp.

XX FR2758331-A1.

XX 17-JUL-1998.

XX 14-JAN-1997; 97FR-0000300.

XX 14-JAN-1997; 97FR-0000300.

XX (UYBO-) UNIV BOURGOGNE.

XX Bourgeois C, Kohli E, Pothier P;

XX WPI: 1998-390320/34.

XX New peptide(s) recognising viral epitope with tropism to mucosa -  
PT useful for, e.g. diagnosing, preventing and treating viral  
PT infection(s)

XX Disclosure; Fig 2; 51pp; French.

XX The present sequence represents the light chain of monoclonal antibody  
CC RS-348, which is directed against the respiratory syncytial virus (RSV).

XX The specification describes peptides which recognise, by

XX antigen-antibody type reactions, at least 1 epitope of a pathogenic virus

XX having tropism for the mucosa. AAW70903-16 and AAW70923-46 are analogous

XX to CDR regions of monoclonal antibodies specific for RSV. AAW70917-28 are

XX analogous to CDR regions of monoclonal antibodies specific for site III

XX or IV of the VP6 protein of rota virus (RV). The peptides can neutralise

XX viral infections and may also inhibit fusion between infected and

XX uninfected cells or cells and viruses. They provide passive or active

XX protection and/or inhibit transcription of the virus, so are useful as

XX antiviral agents or for prophylaxis, in human or veterinary medicine. The

XX peptides can be labelled and used to diagnose infection or contamination

XX by the virus. The peptides are particularly directed against RSV or RS

XX but may also be used against papilloma, adeno, entero, polio, influenza

XX or immune deficiency viruses.

XX Sequence 100 AA;

Query Match 100.0%; Score 46; DB 19; Length 100;

Best Local Similarity 100.0%; Pred. No. 0.25; 0; Indels 0; Gaps 0;

OY 1 PKLLIYWA 8

DB 39 pkllywa 46

## RESULT 6

AAW31647  
ID AAW31647 standard; Protein; 106 AA.

AAW31647:

21-MAY-1998 (first entry)

Monoclonal antibody CP.B8 light chain variable region.

XX  
CYtokine receptor; gamma common chain; gc chain; human;  
KM blocking agent; monoclonal antibody; CP.B8; immunological disease;  
KM myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;  
KM insulin-dependent diabetes; inflammatory bowel disease;  
KM sympathetic ophthalmia; uveitis; allergy; asthma; infection;  
KM graft versus host disease; psoriasis; immunosuppressive; therapy;  
KM complementarity determining region; CDR.

XX Mus musculus.

XX Key Location/Qualifiers

XX Region 24..34 /note= "CDR1"

XX Region 50..56 /note= "CDR2"

XX Region 89..97 /note= "CDR3"

XX MO9743416-A1.

XX 20-NOV-1997.

XX 09-MAY-1997; 97MO-US07870.

XX 10-MAY-1996; 96US-0017466.

XX (BIOJ ) BIOGEN INC.

XX Benjamin CD, Burkly LC, Hession C, Whitty A;

XX WPI: 1998-008885/01.

XX N-PSDB; AAT97440.

XX Blocking agents of the gamma common chain of cytokine receptors -  
XX particularly monoclonal antibodies, used to induce T cell anergy for  
XX treatment of immunological diseases

XX Claim 22; Page 81; 111pp; English.

XX This polypeptide comprises the light chain variable region (VL) of  
XX monoclonal antibody (Mab) C9.B8, which is produced by a hybridoma  
XX deposited as ATCC 12107, and which is specific for the gamma  
XX constant (gc) chain (see AAW31646) of human cytokine receptors. The  
XX invention provides compositions and methods for inhibiting cytokine  
XX signaling using gc chain blocking agents for the treatment of  
XX immunological diseases such as myasthenia gravis, rheumatoid  
XX arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,  
XX inflammatory bowel disease, sympathetic ophthalmia, uveitis,  
XX allergy, asthma, parasitic infection, graft vs. host disease or  
XX psoriasis. Preferred gc blocking agents include Mab CP.B8, its Fab  
XX fragment and an antibody having a light chain variable region  
XX CDR selected from those of CP.B8 VL or a heavy chain variable  
XX region CDR selected from those of CP.B8 VH (see AAW31648).

XX Sequence 106 AA;

Query Match 100.0%; Score 46; DB 19; Length 106;  
Best Local Similarity 100.0%; Pred. No. 0.27;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PKLIYWA 8  
DB 44 pkllywa 51

## RESULT 7

AAK25723  
ID AAK25723 standard; Protein; 107 AA.

AAK25723:

13-JAN-1993 (first entry)

Humanised VL region of the mouse Fd 138-80 antibody.

XX Murine; immunoglobulin; CDR; HSV; non immunogenic; herpes simplex  
XX virus; HSV; Light chain; variable region; framework; human; Eu.  
XX Mus musculus.

XX Key Location/Qualifiers

XX Region 24..38 /note= "CDR"

XX Region 50..56 /note= "CDR"

XX Region 89..97 /note= "CDR"

XX Misc-difference 36 /note= "mutated residue"

XX Misc-difference 48 /note= "mutated residue"

XX Misc-difference 63 /note= "mutated residue"

XX Misc-difference 87 /note= "mutated residue"

XX WO9211018-A.

XX 09-JUL-1992.

XX 19-DEC-1991; 91MO-US09711.

XX 19-DEC-1990; 90US-0634278.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX Co MS, Coeligh KL, Landolfi NF, Queen CL, Schneider WP;

XX WPI: 1992-249842/30.

XX New immunoglobulin(s) having murine CDRs in human framework  
XX regions - have lower antigenicity; useful for treating e.g. HSV,  
XX CMV, T-cell disorders, myeloid disorders and auto-immune  
XX conditions

XX Claim 19; Fig 14D; 141pp; English.

XX The sequence shows the humanised mature light chain variable region  
XX of the mouse Fd 138-80 antibody. Murine CDRs were used in a  
XX human Eu framework to produce a pure humanised immunoglobulin (Ig)  
XX which is capable of binding to a herpes simplex virus-specific  
XX epitope. The Ig is non immunogenic, due to the human framework,  
XX and has a strong affinity for its predetermined antigen. They can  
XX be produced in large quantities via recombinant DNA and monoclonal  
XX antibody technology. The humanised Igs may be used alone or in  
XX combination with chemotherapeutic agents such as non-steroidal  
XX anti-inflammatory drugs or immunosuppressants.  
XX See also AAK25721-32.

XX Sequence 107 AA;

Query Match 100.0%; Score 46; DB 13; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 0.27;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLIYWA 8  
 |||||  
 DB 44 PKLIYWA 51

## RESULT 8

AAAR38599 ID AAR38599 standard; peptide; 107 AA.

AC AAR38599;

DE 28-OCT-1993 (first entry)

DE Human heavy chain subgroup 1 (hh1).

KW Antibody; variable domain; light; L; heavy; H; consensus;  
 affinity; antigen; immunogenicity; humanisation; framework.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 96 /note- "residue conserved in less than 50% of the  
 known sequences of hh1"

FT Misc-difference 104 /note- "residue conserved in less than 50% of the  
 known sequences of hh1"

FT W09311794-A.

PN 24-JUN-1993.

PD 14-DEC-1992; 92MO-US10906.

PF 13-DEC-1991; 91US-0808464.

XX (XOMA ) XOMA CORP.

XX Fishwild DM, Kohn FR, Little RG, Studnicka GM;

XX WPI; 1993-213827/26.

DR Antibodies prep. used for treatment of auto-immune diseases - by  
 replacement of critical residues to reduce immunogenicity but  
 retain binding affinity, etc.

XX Claim 2; Page 98-99; 160pp; English.

XX The consensus amino acid sequences for the subgroups of light  
 chains (hk1 - AAR38590, hk3 - NGK, hk2 - GST, hL1 - AAR38591, hL2 -  
 AAR38592, hL3 - AAR38593, hL6 - AAR38594, hK4 - AAR38595, hL4 -  
 AAR38596, and hL5 - AAR38597) and heavy chains (hh3 - AAR38598, hhl -  
 AAR38599, and hh2 - AAR38600) of human variable domains may be used to  
 prepare, for example, a modified mouse antibody variable domain that  
 retains the affinity of the natural domain for antigen while exhibiting  
 reduced immunogenicity in humans.

XX Unlike other methods of humanisation, which advocate the  
 replacement of entire antibody framework regions with those of human  
 antibodies, this method involves only the introduction of human  
 residues into those positions not critical for antigen binding.

XX This ensures that the binding properties of the modified antibody  
 are not diminished.

XX Sequence 107 AA;

XX Query Match 100.0%; Score 46; DB 14; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 0.27;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PKLIYWA 8  
 |||||  
 DB 45 PKLIYWA 52

## RESULT 9

AAW80817 ID AAW80817 standard; Protein; 107 AA.

AC AAW80817;

DE 16-FEB-1999 (first entry)

DE Amino acid sequence of human kappa 5.1 light chain variable region.

KW Human; kappa 5.1 light chain variable region; receptor; antigen;  
 tumour; auto-immune disease; graft rejection; allergy;

KW Inflammatory disease; endocrine disease; degenerative disease.

OS Homo sapiens.

XX W09846645-A2.

XX 22-OCT-1998.

XX 14-APR-1998; 98MO-EP02180.

XX 14-APR-1997; 97EP-0106109.

XX (KUFE/) KUFE P.

XX (RAUM/) RAUM T.

XX Kufer P, Raum T;

XX WPI; 1998-594564/50.

XX N-PSDB; AAW68539.

XX Production of anti-human antigen receptors - by selecting a  
 combination of functionally rearranged VH and VL immunoglobulin  
 chains expressed from a recombinant vector

XX Claim 9; Fig 9; 84pp; English.

XX This is the amino acid sequence of the human kappa light chain  
 variable region, used in the method of the invention, for providing  
 receptors that can be used for targeting antigens in humans without  
 being immunogenic themselves. Such receptors can be used for treating  
 diseases such as tumours or auto-immune diseases, graft rejection  
 after transplantation, infectious diseases by targeting cellular  
 receptors as well as allergic, inflammatory, endocrine and  
 degenerative diseases by targeting key molecules involved in the  
 pathological process.

XX Sequence 107 AA;

Query Match 100.0%; Score 46; DB 19; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 0.27;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLIYWA 8  
 |||||  
 DB 44 PKLIYWA 51

## RESULT 10

AAW80814 ID AAW80814 standard; Protein; 107 AA.

AC AAW80814;

DT 16-FEB-1999 (first entry)  
 XX Amino acid sequence of human Kappa 8 light chain variable region.  
 DE  
 XX  
 KW Human; kappa 8 light chain variable region; receptor; antigen;  
 KW tumour; auto-immune disease; graft rejection; allergy;  
 KW inflammatory disease; endocrine disease; degenerative disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9846645-A2.  
 XX  
 XX 22-OCT-1998.  
 PD  
 XX 14-APR-1998; 98WO-EP02180.  
 PF  
 XX 14-APR-1997; 97EP-0106109.  
 PR  
 XX (KUFE/) KUFE P.  
 PA (RAUM/) RAUM T.  
 XX  
 PI Kufer P, Raum T;  
 XX  
 DR WPI: 1998-594564/50.  
 DR N-PSDB; AAV68536.  
 XX  
 PT Production of anti-human antigen receptors - by selecting a  
 PT combinatorial library of functionally rearranged VH and VL immunoglobulin  
 PT chains expressed from a recombinant vector  
 XX  
 PS Claim 9; Fig 6; 84pp; English.  
 XX  
 CC This is the protein sequence of the human kappa 8 light chain  
 CC variable region, used in the method of the invention, for providing  
 CC receptors that can be used for targeting antigens in humans without  
 CC being immunogenic themselves. Such receptors can be used for treating  
 CC diseases such as tumours or auto-immune diseases, graft rejection  
 CC after transplantation, infectious diseases, by targeting cellular  
 CC receptors as well as allergic, inflammatory, endocrine and  
 CC degenerative diseases by targeting key molecules involved in the  
 CC pathological processes.  
 CC  
 XX Sequence 107 AA;  
 SO

Query Match 100.0%; Score 46; DB 19; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 0.27;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLIYWA 8  
 DB 44 PKLIYWA 51

RESULT 11  
 AAM58499  
 ID AAM58499 standard; protein; 107 AA.  
 XX  
 AC AAM58499;  
 XX  
 DT 18-AUG-1998 (first entry)  
 DE Human kappa light chain subgroup 4 consensus sequence hK4.  
 XX  
 KW Humanised; human; mouse; CD5; anti-CD5 antibody; immunoglobulin;  
 KW depletion; cytotoxic; immunocongulate; fusion protein; psoriasis;  
 KW autoimmune disease; rheumatoid arthritis; type I diabetes.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 96

FT /note="not specified"  
 FT Misc-difference 104  
 FT /note="not specified"  
 XX  
 PN US5770196-A.  
 XX  
 PD 23-JUN-1998.  
 XX  
 PF 07-JUN-1995; 95US-0472788.  
 PR  
 XX 23-JUN-1993; 93US-0082842.  
 PR 13-DEC-1991; 91US-0808464.  
 PR 14-DEC-1992; 92WO-US10906.  
 PR 07-JUN-1995; 95US-0472788.  
 XX  
 PA (XOMA ) XOMA CORP.  
 XX  
 PI Studnicka GW;  
 XX  
 DR WPI: 1998-376744/32.  
 XX  
 PT Depletion of CD5-positive cells in vivo - using anti-CD5 antibodies  
 PT with humanised variable regions  
 XX  
 PS Disclosure; Column 57-58; 77pp; English.  
 XX  
 CC A method has been developed of depleting CD5+ cells in an animal. The  
 CC method comprises administering a cytotoxic protein containing a modified  
 CC immunoglobulin (Ig) variable domain, where the protein is an anti-CD5 Ig  
 CC molecule or an immunocongulate or fusion protein containing an anti-CD5  
 CC Ig molecule, and where the modified Ig variable domain comprises at  
 CC least one of (a) a modified light chain variable region (see AAM58479  
 CC or AAM58480), and (b) a modified heavy chain variable region (see AAM58479  
 CC or AAM58481), where AAM58478 and AAM58479 are humanised forms of the H65  
 CC light and heavy chain variable domains with low risk amino acid  
 CC substitutions [i.e. low risk of reducing antigen-binding specificity.]  
 CC and AAM58480 and AAM58481 are humanised forms of the H65 light and heavy  
 CC chain variable domains with moderate risk amino acid substitutions and  
 CC are present in humanised H65 antibody h63 (ATCC HB 11206). The method is  
 CC useful for treating autoimmune diseases, especially systemic lupus  
 CC erythematosus, rheumatoid arthritis, psoriasis or type I diabetes. The  
 CC present sequence represents a consensus amino acid sequence for light  
 CC chain subgroups of human antibody variable domains, from the present  
 CC invention.  
 CC  
 XX Sequence 107 AA;  
 SO

Query Match 100.0%; Score 46; DB 19; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 0.27;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLIYWA 8  
 DB 45 PKLIYWA 52

RESULT 12  
 AAM47087  
 ID AAM47087 standard; Protein; 107 AA.  
 XX  
 AC AAM47087;  
 XX  
 DT 26-JUN-1998 (first entry)  
 DE Mouse J591 monoclonal antibody light chain variable region VK17.  
 XX  
 KW Mouse; monoclonal antibody; J591; prostate specific membrane antigen;  
 KW cancer; vascular endothelial cell; metastatic adenocarcinoma.  
 XX  
 OS Mus sp.  
 XX  
 PN WO9803873-A1.



```

XX          29-JAN-1998.
XX
XX          17-JUL-1997;      97WO-US12035.
XX
XX          09-APR-1997;      97US-0838682.
XX
XX          18-JUL-1996;      96US-0022125.
XX
XX          (CORR ) CORNELL RES FOUND INC.
XX
XX          Bander NH;
XX
XX          WPI; 1998-120937/11.
XX
XX          DR          N-PSDB; AAV13953.
XX
XX          Destroying cancer cells with agent that binds to prostate specific
XX          membrane antigen - on vascular endothelial cells near the cancer, or
XX          on normal, hypertrophic or cancerous prostatic cells, also used for
XX          diagnosis
XX
XX          Example 12; Page 61; 94pp; English.
XX
XX          The present sequence represents the mouse J591 monoclonal antibody light
XX          chain variable region from clone VX17 from an example of the present
XX          invention. The present invention describes the elimination of cancer
XX          cells by treating vascular endothelial cells (VEC) close to the cancer
XX          with an agent (A) able to bind to the extracellular domain (ECD) of
XX          prostate specific membrane antigen (PSMA). (A) both binds to the VEC and
XX          destroys the cancer cells. Also described are: (1) the detection of
XX          cancer tissue by detecting binding of labelled (A) to VEC close to, or
XX          within, a cancer tissue; (2) eliminating or detecting normal, benignly
XX          hyperplastic or cancerous prostate epithelial cells using optionally
XX          labelled (A); (3) hybridomas that produce a monoclonal antibody (Mab)
XX          that binds to PSMA. The method is used to treat renal, urothelial,
XX          colon, lung, rectal or breast cancers and metastatic adenocarcinoma of
XX          the liver. The diagnostic method is particularly used to detect
XX          recurrence of prostatic disease or to monitor the effect of treatments
XX          for prostate cancer (presence of PSMA in the serum indicates that
XX          prostate cells are being lysed). (A) binds to an epitope of PSMA
XX          expressed on alive cells (contrast antibody 7E11 which only binds after
XX          cell lysis), allowing targeting of live, unfixed cells and thus
XX          providing more efficient treatment and diagnosis. Both cancer cells
XX          themselves and the VEC on which they depend are killed. All VEC close to
XX          cancer cells express PSMA, whatever the type of cancer, but normal VEC
XX          do not.
XX
XX          Sequence      107 AA;
XX
XX          Query Match          100.0%; Score 46; DB 19; Length 107;
XX          Best Local Similarity 100.0%; Pred. No. 0 27;
XX          Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0.
XX
XX          1 PKLLITWA 8
XX          |||||||
XX          44 PKLLITWA 51
XX
XX          Db
XX
XX          RESULT 13
XX          AAV17955
XX          ID          AAV17955 standard; Protein: 107 AA.
XX
XX          AAV17955;
XX
XX          04-AUG-1999 (first entry)
XX
XX          Human kappa 8 light chain variable region.
XX
XX          Binding site domain: BSD; epitope: fusion protein: therapeutic; cancer;
XX          autoimmune disease; scFv-antibody; single-chain Fv.
XX
XX          Homo sapiens.
XX

```

PN	MO9925818-AL.
XX	27-MAY-1999.
PD	
PF	16-NOV-1998; 98MO-EP07313.
XX	
PR	17-NOV-1997; 97EP-0120096.
XX	
PA	(KUFE/) KUFER P.
XX	
PI	Borschett K, Kufer P, Lutterbuese R, Raum T, Zetzl F;
DR	WPI; 1999-338004/28.
N-PSDB;	AAX77237.
XX	
PT	Phase display system for identification of binding site domains
PT	retaining capacity to bind an epitope
XX	
PS	Disclosure; Fig 3.2; 152pp; English.
CC	The invention relates to a method of identifying binding site domains
CC	(BSD) that retain the capacity of binding to a predetermined epitope when
CC	positioned C-terminal of at least one further domain in a recombinant bi-
CC	or multivalent polypeptide. The method comprises (a) testing a panel of
CC	BSD displayed on the surface of a biological display system as part of a
CC	fusion protein for binding to a predetermined epitope, where the fusion
CC	protein comprises an additional domain positioned N-terminal of the BSD
CC	and an amino acid sequence that mediates anchoring of the fusion protein
CC	to the surface of the display system; and (b) identifying a BSD that
CC	binds to the predetermined epitope. The method is useful to identify bi-
CC	or multivalent polypeptides that comprise antibody binding sites capable
CC	of efficiently binding to the corresponding antigen. The polypeptides or
CC	antibodies identified by the method are useful therapeutically and
CC	diagnostically, for e.g. cancer and autoimmune diseases. scFv-antibody
CC	fragments that bind independently of their position within bifunctional
CC	single-chain fusion proteins can be isolated from combinatorial antibody
CC	libraries using the new in vitro method.
SO	
Sequence	107 AA;
Query Match	100.0%; Score 46; DB 20; Length 107;
Best Local Similarity	100.0%; Pred. No. 0.27;
Matches	8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Gy	1 PKLLIYWA 8       
Db	44 pKLIlywa 51
RESULT 14	
ID	AAY90375
AC	AAY90375 standard; Protein; 107 AA.
XX	
AA	AAY90375;
XX	
DT	15-JAN-2001 (first entry)
XX	
DE	J591 monoclonal antibody light chain protein sequence fragment.
XX	
KW	J591 monoclonal antibody; extracellular domain; diagnosis; therapy;
KW	prostate specific membrane antigen; prostate cancer; light chain.
OS	Homo sapiens.
PN	US6107090-A.
XX	
PD	22-AUG-2000.
XX	
PF	09-APR-1997; 97US-0838682.
XX	
PR	06-MAY-1996; 96US-0016976.
PR	18-JUL-1996; 96US-0022125.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 12:59:17 ; Search time 75.67 Seconds  
(without alignments)  
2.582 Million cell updates/sec

Title: US-09-712-819A-4

Perfect score: 46

Sequence: 1 PRLTYMA 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/prodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCrUS.COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	100.0	32	3 US-08-513-968-55	Sequence 55, Appl
2	46	100.0	107	1 US-07-634-278-50	Sequence 50, Appl
3	46	100.0	107	1 US-07-634-278-51	Sequence 51, Appl
4	46	100.0	107	1 US-08-477-728-50	Sequence 50, Appl
5	46	100.0	107	1 US-08-477-728-51	Sequence 51, Appl
6	46	100.0	107	1 US-08-474-040-50	Sequence 50, Appl
7	46	100.0	107	1 US-08-474-040-51	Sequence 51, Appl
8	46	100.0	107	1 US-08-487-200-50	Sequence 50, Appl
9	46	100.0	107	1 US-08-487-200-51	Sequence 51, Appl
10	46	100.0	107	1 US-08-488-113B-156	Sequence 156, App
11	46	100.0	107	1 US-08-477-484B-156	Sequence 156, App
12	46	100.0	107	1 US-08-107-669P-20	Sequence 20, Appl
13	46	100.0	107	2 US-08-472-788A-20	Sequence 20, Appl
14	46	100.0	107	2 US-08-477-531B-20	Sequence 20, Appl
15	46	100.0	107	2 US-08-646-360-156	Sequence 156, App
16	46	100.0	107	2 US-08-082-842A-20	Sequence 20, Appl
17	46	100.0	107	3 US-08-838-682-19	Sequence 19, Appl
18	46	100.0	107	4 US-08-895-914-19	Sequence 19, Appl
19	46	100.0	107	4 US-08-839-765-156	Sequence 156, App
20	46	100.0	107	4 US-09-136-389-156	Sequence 156, App
21	46	100.0	107	4 US-08-484-537-50	Sequence 50, Appl
22	46	100.0	107	4 US-08-464-537-51	Sequence 51, Appl
23	46	100.0	107	4 US-09-357-710A-19	Sequence 19, Appl
24	46	100.0	109	2 US-08-308-494A-23	Sequence 23, Appl
25	46	100.0	110	4 US-08-957-001B-5	Sequence 5, Appl
26	46	100.0	110	4 US-08-957-001B-24	Sequence 24, Appl
27	46	100.0	110	4 US-09-496-301-5	Sequence 5, Appl

28	46	100.0	110	4 US-09-496-301-24	Sequence 24, Appl
29	46	100.0	112	2 US-07-942-245-30	Sequence 30, Appl
30	46	100.0	112	2 US-07-916-098A-15	Sequence 15, Appl
31	46	100.0	113	1 US-08-690-102A-2	Sequence 2, Appl
32	46	100.0	113	1 US-08-690-102A-6	Sequence 6, Appl
33	46	100.0	113	2 US-08-263-911-2	Sequence 2, Appl
34	46	100.0	113	2 US-08-819-033-1	Sequence 1, Appl
35	46	100.0	113	3 US-08-483-749A-16	Sequence 16, Appl
36	46	100.0	113	3 US-08-513-968-82	Sequence 82, Appl
37	46	100.0	113	4 US-09-127-902-2	Sequence 2, Appl
38	46	100.0	113	4 US-09-127-902-6	Sequence 6, Appl
39	46	100.0	113	4 US-09-155-107-2	Sequence 2, Appl
40	46	100.0	113	4 US-09-155-107-6	Sequence 6, Appl
41	46	100.0	113	4 US-09-155-107-20	Sequence 20, Appl
42	46	100.0	113	4 US-08-525-539A-80	Sequence 80, Appl
43	46	100.0	113	4 US-09-025-203-7	Sequence 7, Appl
44	46	100.0	113	4 US-09-025-203-8	Sequence 8, Appl
45	46	100.0	113	4 US-09-025-203-9	Sequence 9, Appl

## ALIGNMENTS

RESULT 1  
US-08-513-968-55

Sequence 55 Application US/08513968

Patent No. 611413

GENERAL INFORMATION:

APPLICANT: EDA, Yasuyuki

APPLICANT: MAEDA, Hiroaki

APPLICANT: MAKIZUMI, Keiichi

APPLICANT: SHIOSAKI, Kouichi

APPLICANT: OSATOMI, Kazuhiko

APPLICANT: KIMACHI, Hirofumi

APPLICANT: TOKIYOSHI, Seichiro

TITLE OF INVENTION: ANTI-HIV MONOCLONAL ANTIBODY

NUMBER OF SEQUENCES: 86

CORRESPONDENCE ADDRESSES:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

FILING DATE: 11-SEP-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 78913/1993

FILING DATE: 11-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: EDA-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

TELEX: 248633

INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 32 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-513-968-55

Query Match 100.0%; Score 46; DB 3; Length 32;  
Best Local Similarity 100.0%; Pred. No. 0.04; Mismatches 8; Conservative 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLLIYMA 8  
| | | | | | | |  
DB 8 PKLLIYMA 15

RESULT 2  
US-07-634-278-50  
; Sequence 50, Application US/07634278  
; Patent No. 5530101  
; GENERAL INFORMATION:  
; APPLICANT: QUEEN, Cary L.  
; APPLICANT: CO, Man Sung  
; APPLICANT: SCHNEIDER, William P.  
; APPLICANT: LANDOLFI, Nicholas F.  
; APPLICANT: COELINGH, Kathleen L.  
; APPLICANT: SELICK, Harold E.  
; TITLE OF INVENTION: IMPROVED HUMANIZED IMMUNOGLOBULINS  
; NUMBER OF SEQUENCES: 113  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourile and Crew  
; STREET: 379 Lytton Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: US  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 13-FEB-1989  
; APPLICATION NUMBER: US 07/310,252  
; PRIORITY APPLICATION DATA:  
; FILING DATE: 19-DEC-1990  
; CLASSIFICATION: 424  
; PRIOR APPLICATION NUMBER: US 07/590,274  
; FILING DATE: 28-SEP-1990  
; APPLICATION DATA:  
; FILING DATE: 13-FEB-1989  
; APPLICATION NUMBER: US 07/290,975  
; FILING DATE: 28-DEC-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, William M  
; REGISTRATION NUMBER: 30,223  
; REFERENCE/DOCKET NUMBER: 11823-002600  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 326-2400  
; TELEFAX: (415) 326-2422  
; INFORMATION FOR SEQ ID NO: 50:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 107 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-07-634-278-50

Query Match 100.0%; Score 46; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLLIYMA 8  
| | | | | | | |  
DB 44 PKLLIYMA 51

RESULT 3  
US-07-634-278-51  
; Sequence 51, Application US/07634278  
; Patent No. 5530101  
; GENERAL INFORMATION:  
; APPLICANT: QUEEN, Cary L.  
; APPLICANT: CO, Man Sung  
; APPLICANT: SCHNEIDER, William P.  
; APPLICANT: LANDOLFI, Nicholas F.  
; APPLICANT: COELINGH, Kathleen L.  
; APPLICANT: SELICK, Harold E.  
; TITLE OF INVENTION: IMPROVED HUMANIZED IMMUNOGLOBULINS  
; NUMBER OF SEQUENCES: 113  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourile and Crew  
; STREET: 379 Lytton Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: US  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 13-FEB-1989  
; APPLICATION NUMBER: US 07/310,252  
; PRIORITY APPLICATION DATA:  
; FILING DATE: 19-DEC-1990  
; CLASSIFICATION: 424  
; PRIOR APPLICATION NUMBER: US 07/590,274  
; FILING DATE: 28-SEP-1990  
; APPLICATION DATA:  
; FILING DATE: 13-FEB-1989  
; APPLICATION NUMBER: US 07/290,975  
; FILING DATE: 28-DEC-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, William M  
; REGISTRATION NUMBER: 30,223  
; REFERENCE/DOCKET NUMBER: 11823-002600  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 326-2400  
; TELEFAX: (415) 326-2422  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 107 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-07-634-278-51

Query Match 100.0%; Score 46; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLLIYMA 8  
| | | | | | | |  
DB 44 PKLLIYMA 51

RESULT 4  
US-08-477-728-50  
; Sequence 50, Application US/08477728  
; Patent No. 5565089  
; GENERAL INFORMATION:  
; APPLICANT: QUEEN, Cary L.  
; APPLICANT: SCHNEIDER, William P.  
; APPLICANT: SELICK, Harold E.

TITLE OF INVENTION: IMPROVED HUMANIZED IMMUNOGLOBULINS  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: Palo Alto  
STATE: California  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,728  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/634,278  
FILING DATE: 19-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/590,274  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/310,252  
FILING DATE: 13-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/290,975  
FILING DATE: 28-DEC-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 11823-002600  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 107 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-477-728-50  
Query Match 100.0%; Score 46; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLIYWA 8  
|||||||  
Db 44 PKLIYWA 51

RESULT 5  
US-08-477-728-51  
Sequence 51, Application US/08477728  
Patent No. 5585089  
GENERAL INFORMATION:  
APPLICANT: QUEEN, Cary L.  
APPLICANT: SCHNEIDER, William P.  
APPLICANT: SELICK, Harold E.  
TITLE OF INVENTION: IMPROVED HUMANIZED IMMUNOGLOBULINS  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: Palo Alto  
STATE: California  
COUNTRY: US  
ZIP: 94111

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,728  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/634,278  
FILING DATE: 19-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/590,274  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/310,252  
FILING DATE: 13-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/290,975  
FILING DATE: 28-DEC-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 11823-002600  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 107 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-477-728-51  
Query Match 100.0%; Score 46; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLIYWA 8  
|||||||  
Db 44 PKLIYWA 51

RESULT 6  
US-08-474-040-50  
Sequence 50, Application US/08474040  
Patent No. 563761  
GENERAL INFORMATION:  
APPLICANT: QUEEN, Cary L.  
APPLICANT: CO, Man Sung  
APPLICANT: SCHNEIDER, William P.  
APPLICANT: LANDOLFI, Nicholas F.  
APPLICANT: COELINGH, Kathleen L.  
APPLICANT: SELICK, Harold E.  
TITLE OF INVENTION: IMPROVED HUMANIZED IMMUNOGLOBULINS  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: 379 Lytton Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: US  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/474,040  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/634,278  
FILING DATE: 19-DEC-1990  
APPLICATION NUMBER: US 07/590,274  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/310,252  
FILING DATE: 13-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/290,975  
FILING DATE: 28-DEC-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: SMITH, WILLIAM M  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 11823-002600  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 107 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-474-040-50

Query Match 100.0%; Score 46; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLIYWA 8  
|||||||  
DB 44 PKLIYWA 51

RESULT 7  
US-08-474-040-51  
Sequence 51, Application US/08474040  
Patent No. 5693761  
GENERAL INFORMATION:  
APPLICANT: QUEEN, Cary L.  
APPLICANT: CO, Man Sung  
APPLICANT: SCHNEIDER, William P.  
APPLICANT: LANDOLFI, Nicholas F.  
APPLICANT: COELINGH, Kathleen L.  
APPLICANT: SELICK, Harold E.  
TITLE OF INVENTION: IMPROVED HUMANIZED IMMUNOGLOBULINS  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourile and Crew  
STREET: 379 Lytton Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: US  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,040  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/634,278  
FILING DATE: 19-DEC-1990  
APPLICATION NUMBER: US 07/590,274

FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/310,252  
FILING DATE: 13-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/290,975  
FILING DATE: 28-DEC-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: SMITH, WILLIAM M  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 11823-002600  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 107 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-474-040-51

Query Match 100.0%; Score 46; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLIYWA 8  
|||||||  
DB 44 PKLIYWA 51

RESULT 8  
US-08-487-200-50  
Sequence 50, Application US/08487200  
Patent No. 5693762  
GENERAL INFORMATION:  
APPLICANT: QUEEN, Cary L.  
APPLICANT: CO, Man Sung  
APPLICANT: SCHNEIDER, William P.  
APPLICANT: LANDOLFI, Nicholas F.  
APPLICANT: COELINGH, Kathleen L.  
APPLICANT: SELICK, Harold E.  
TITLE OF INVENTION: IMPROVED HUMANIZED IMMUNOGLOBULINS  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: 379 Lytton Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: US  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,200  
FILING DATE: 7-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/634,278  
FILING DATE: 19-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/590,274  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/310,252  
FILING DATE: 13-FEB-1989  
APPLICATION NUMBER: US 07/290,975

FILING DATE: 28-DEC-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 11823-002610  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 107 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-487-200-50

Query Match 100.0%; Score 46; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLITYA 8  
Db 44 PKLITYA 51

RESULT 9  
US-08-487-200-51  
Sequence 51, Application US/08487200  
Patent No. 5693762  
GENERAL INFORMATION:  
APPLICANT: QUEEN, Cary L.  
APPLICANT: CO. Man Sung  
APPLICANT: SCHNEIDER, William P.  
APPLICANT: LANDOLFI, Nicholas F.  
APPLICANT: COBLEIGH, Kathleen L.  
APPLICANT: SELICK, Harold E.  
TITLE OF INVENTION: IMPROVED HUMANIZED IMMUNOGLOBULINS  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Townsend and Townsend and Crew  
STREET: 379 Lytton Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: US  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,200  
FILING DATE: 7-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/654,278  
FILING DATE: 19-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/590,274  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/310,252  
FILING DATE: 13-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/290,975  
FILING DATE: 28-DEC-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William M  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 11823-002610  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 107 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-487-200-51

Query Match 100.0%; Score 46; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLITYA 8  
Db 44 PKLITYA 51

RESULT 10  
US-08-488-113B-156  
Sequence 156, Application US/08488113B  
Patent No. 5744580  
GENERAL INFORMATION:  
APPLICANT: Better, Marc D.  
APPLICANT: Carroll, Stephen F.  
APPLICANT: Studulka, Gary M.  
TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating  
NUMBER OF SEQUENCES: 169  
CORRESPONDENCE ADDRESS:  
ADDRESSER: McAndrews, Held & Malloy, Ltd.  
STREET: 500 West Madison Street, 34th floor  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60661  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,113B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/425,336  
FILING DATE: 18-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/064,691  
FILING DATE: 12-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/988,430  
FILING DATE: 09-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/901,707  
FILING DATE: 19-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/787,567  
FILING DATE: 04-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: McNicholas, Janet M.  
REGISTRATION NUMBER: 32,918  
REFERENCE/DOCKET NUMBER: 11022US07/200-70.P3.C2A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/707-8889  
TELEFAX: 312/707-9155  
TELEX: 650 388-1248  
INFORMATION FOR SEQ ID NO: 156:  
SEQUENCE CHARACTERISTICS:

LENGTH: 107 amino acids  
TYPE: amino acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-488-113B-156

Query Match 100.0%; Score 46; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLIYWA 8  
Db 45 PKLIYWA 52

RESULT 11  
US-08-477-484B-156  
Sequence 156, Application US/08477484B  
Patent No. 5756699  
GENERAL INFORMATION:  
APPLICANT: Belter, Marc D.  
APPLICANT: Carroll, Stephen F.  
APPLICANT: Studnicka, Gary M.  
TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating  
NUMBER OF SEQUENCES: 169  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McAndrews, Held & Malloy, Ltd.  
STREET: 500 West Madison Street, 34th Floor  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60661  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,484B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/425,336  
FILING DATE: 18-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/064,691  
FILING DATE: 12-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/988,430  
FILING DATE: 09-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/901,707  
FILING DATE: 19-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/787,567  
FILING DATE: 04-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: McNicholas, Janet M.  
REGISTRATION NUMBER: 32,918  
REFERENCE/DOCKET NUMBER: 1102US07/200-70.P3.C2A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/707-8889  
TELEFAX: 312/707-9155  
TELEX: 650 388-1248  
INFORMATION FOR SEQ ID NO: 156:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 107 amino acids  
TYPE: amino acid  
STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-477-484B-156

Query Match 100.0%; Score 46; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLIYWA 8  
Db 45 PKLIYWA 52

RESULT 12  
US-08-107-669D-20  
Sequence 20, Application US/08107669D  
Patent No. 5766866  
GENERAL INFORMATION:  
APPLICANT: Studnicka, Gary M.  
TITLE OF INVENTION: Modified Antibody Variable Domains (as amended)  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein and Fox P.L.L.C.  
STREET: 1100 New York Ave., N.W., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/107,669D  
FILING DATE: 13-AUG-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/10906  
FILING DATE: 14-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/808,464  
FILING DATE: 13-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Michele A. Cimbara  
REGISTRATION NUMBER: 33,851  
REFERENCE/DOCKET NUMBER: 0610.1000001/MAC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202/371-2540  
TELEFAX: 202/371-2540  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 107 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-107-669D-20

Query Match 100.0%; Score 46; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PKLIYWA 8  
Db 45 PKLIYWA 52

RESULT 13  
US-08-472-788A-20  
Sequence 20, Application US/08472788A  
Patent No. 5770196  
GENERAL INFORMATION:



APPLICANT: Studnicka, Gary M.  
TITLE OF INVENTION: Modified Antibody Variable Domains  
NUMBER OF SEQUENCES: 89  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein and Fox P.L.L.C.  
STREET: 1100 New York Ave., N.W., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,788A  
FILING DATE: 07-JUN-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/082,842  
FILING DATE: 23-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/10906  
FILING DATE: 14-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/808,464  
FILING DATE: 13-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Cimbala, Michele A.  
REGISTRATION NUMBER: 33,851  
REFERENCE/DOCKET NUMBER: 0610.1000003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202/371-2600  
TELEFAX: 202/371-2540  
TELEX:  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 107 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-472-788A-20

Query Match 100.0%; Score 46; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLIYWA 8  
DB 45 PKLIYWA 52

RESULT 14  
US-08-477-531B-20  
Sequence 20, Application US/08477531B  
Patent No. 5821123  
GENERAL INFORMATION:  
APPLICANT: Studnicka, Gary M.  
TITLE OF INVENTION: Modified Antibody Variable Domains (as amended)  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein and Fox P.L.L.C.  
STREET: 1100 New York Ave., N.W., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,531B  
FILING DATE: 07-JUN-1993  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/107,669  
FILING DATE: 13-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/10906  
FILING DATE: 14-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/808,464  
FILING DATE: 13-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Michele A. Cimbala  
REGISTRATION NUMBER: 33,851  
REFERENCE/DOCKET NUMBER: 0610.1000004/MAC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202/371-2600  
TELEFAX: 202/371-2540  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 107 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-477-531B-20

Query Match 100.0%; Score 46; DB 2; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLIYWA 8  
DB 45 PKLIYWA 52

RESULT 15  
US-08-646-360-156  
Sequence 156, Application US/08646360  
Patent No. 5837491  
GENERAL INFORMATION:  
APPLICANT: Belter, Marc D.  
APPLICANT: Carroll, Stephen F.  
APPLICANT: Studnicka, Gary M.  
TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating  
NUMBER OF SEQUENCES: 173  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McAndrews, Held & Malloy, Ltd.  
STREET: 500 West Madison Street, 34th floor  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60661  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/646,360  
FILING DATE: 13-MAY-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/05348  
FILING DATE: 12-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/064,691  
FILING DATE: 12-MAY-1993  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/988,430  
FILING DATE: 09-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/901,707  
FILING DATE: 19-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/787,567  
FILING DATE: 04-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: MCNICHOLAS, Janet M.  
REGISTRATION NUMBER: 32,918  
REFERENCE/DOCKET NUMBER: 200-70.P4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/707-8889  
TELEFAX: 312/707-9155  
TELEX: 650 388-1248  
INFORMATION FOR SEQ ID NO: 156:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 107 amino acids  
TYPE: amino acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-646-360-156

Query Match 100.0%; Score 46; DB 2; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.13; 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKLITYA 8  
|||||||  
DB 45 PKLITYA 52

Search completed: July 15, 2002, 12:59:18  
Job time: 389 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:01:01 ; Search time 95.45 seconds  
(without alignments)  
7.047 Million cell updates/sec

Title: US-09-712-819a-6

Perfect score: 33

Sequence: 1 FTLKISR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 71:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	54	2 S34093	Ig kappa chain V r
2	33	100.0	71	2 H30538	Ig kappa chain V r
3	33	100.0	72	2 S40358	Ig kappa chain - h
4	33	100.0	75	2 S40337	Ig kappa chain V-J
5	33	100.0	83	2 S34095	Ig kappa chain V r
6	33	100.0	83	2 S24211	Ig kappa chain V r
7	33	100.0	87	2 S34091	Ig kappa chain V r
8	33	100.0	87	2 S34094	Ig kappa chain V r
9	33	100.0	89	2 B25155	Ig kappa chain V r
10	33	100.0	91	2 S25463	Ig kappa chain V r
11	33	100.0	91	2 S42186	Ig kappa chain V r
12	33	100.0	93	2 PH1039	Ig kappa chain V r
13	33	100.0	94	2 PI0258	Ig kappa chain V r
14	33	100.0	96	2 S40320	Ig kappa chain V r
15	33	100.0	99	2 S24501	Ig kappa chain - h
16	33	100.0	99	2 E28833	Ig heavy chain V r
17	33	100.0	99	2 S24504	Ig kappa chain V r
18	33	100.0	100	2 S26334	Ig light chain V r
19	33	100.0	100	2 S24681	Ig kappa chain - h
20	33	100.0	100	2 S45440	Ig kappa chain V r
21	33	100.0	101	2 PH1057	Ig kappa chain V r
22	33	100.0	101	2 S34153	Ig kappa chain V r
23	33	100.0	101	2 A33730	Ig kappa chain V r
24	33	100.0	102	2 PH1035	Ig light chain V r
25	33	100.0	102	2 PH1044	Ig light chain V r
26	33	100.0	102	2 S14592	Ig kappa chain V r
27	33	100.0	102	2 E28195	Ig kappa chain V r
28	33	100.0	102	2 S14594	Ig kappa chain V r
29	33	100.0	102	2 S14593	Ig kappa chain V r

30	33	100.0	102	2 S14591	Ig kappa chain V r
31	33	100.0	102	2 S14590	Ig kappa chain V r
32	33	100.0	103	2 PH1034	Ig light chain V r
33	33	100.0	103	2 PH1036	Ig light chain V r
34	33	100.0	103	2 PH1037	Ig light chain V r
35	33	100.0	103	2 PH1038	Ig light chain V r
36	33	100.0	103	2 PH1040	Ig light chain V r
37	33	100.0	103	2 PH1041	Ig light chain V r
38	33	100.0	103	2 PH1042	Ig light chain V r
39	33	100.0	103	2 PH1043	Ig light chain V r
40	33	100.0	103	2 PH1045	Ig light chain V r
41	33	100.0	103	2 PH1035	Ig light chain V r
42	33	100.0	103	2 PH1056	Ig light chain V r
43	33	100.0	103	2 PH1098	Ig light chain V r
44	33	100.0	103	2 PH1099	Ig light chain V r
45	33	100.0	103	2 PH1100	Ig light chain V r

#### ALIGNMENTS

RESULT 1  
S34093  
Ig kappa chain V region - human  
C:Species: Homo sapiens (man)  
C>Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 07-May-1999  
C:Accession: S34093  
R:Wagner, S.D.; Luzanato, L.  
Eur. J. Immunol. 23, 391-397, 1993  
A:Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distrib  
A:Reference number: S34076; MUID:931170387  
A:Accession: S34093  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 154 <MAG>  
A:Cross-references: EMBL:X67177  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1992  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 33; DB 2; Length 54;  
Best Local Similarity 100.0%; Pred. No. 0.45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLKISR 7  
|||||||  
Db 43 FTLKISR 49

RESULT 2  
H30538  
Ig kappa chain V region (174-394) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 10-Feb-1989 #sequence\_revision 10-Feb-1989 #text\_change 09-May-1997  
C:Accession: H30538  
R:Clafflin, J.L.; Berry, J.  
J. Immunol. 141, 4012-4019, 1988  
A:Title: Genetics of the phosphocholine-specific antibody response to Streptococcus p  
A:Reference number: A30534; MUID:89035545  
A:Accession: H30538  
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual  
A:Molecule type: mRNA  
A:Residues: 1-71 <CLA>  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 33; DB 2; Length 71;  
Best Local Similarity 100.0%; Pred. No. 0.59; 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;

QY 1 FTLKISR 7

Db 35 FTLKISR 41

## RESULT 3

Ig kappa chain - human  
S40358  
C:Species: Homo sapiens (man)  
C>Date: 06-May-1994 #sequence\_revision 26-May-1995 #text\_change 23-Jul-1999  
C:Accession: S40358  
R:Klein, R.; Jaenichen, R.; Zachau, H.G.  
Eur. J. Immunol. 23, 3248-3271, 1993  
A:Title: Expressed human immunoglobulin chl genes and their hypermutation.  
A:Reference number: S40312; MUID:94080891  
A:Accession: S40358  
A:Status: preliminary; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-72 <KLE>  
A:Cross-references: EMBL:X72468; NID:q441404; PIDN:CAA51136.1; PID:q441405  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 33; DB 2; Length 72;  
Best Local Similarity 100.0%; Pred. No. 0.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLKISR 7  
Db 28 FTLKISR 34

## RESULT 4

Ig kappa chain V-J-C region - human  
S40337  
C:Species: Homo sapiens (man)  
C>Date: 19-May-1994 #sequence\_revision 26-May-1995 #text\_change 23-Jul-1999  
C:Accession: S40337  
R:Klein, R.; Jaenichen, R.; Zachau, H.G.  
Eur. J. Immunol. 23, 3248-3271, 1993  
A:Title: Expressed human immunoglobulin chl genes and their hypermutation.  
A:Reference number: S40312; MUID:94080891  
A:Accession: S40337  
A:Status: preliminary; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-75 <KLE>  
A:Cross-references: EMBL:X72447; NID:q441362; PIDN:CAA51115.1; PID:q441363  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 33; DB 2; Length 75;  
Best Local Similarity 100.0%; Pred. No. 0.62;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLKISR 7  
Db 30 FTLKISR 36

## RESULT 5

Ig kappa chain V region - human  
S34095  
C:Species: Homo sapiens (man)  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 21-Jan-2000  
C:Accession: S34095  
R:Wagner, S.D.; Luzzatto, L.  
Eur. J. Immunol. 23, 391-397, 1993  
A:Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distributed  
A:Reference number: S34076; MUID:93170387  
A:Accession: S34095  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-83 <NAG>  
A:Cross-references: EMBL:X67179  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:1-80/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 33; DB 2; Length 83;  
Best Local Similarity 100.0%; Pred. No. 0.68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLKISR 7  
Db 61 FTLKISR 67

## RESULT 6

Ig kappa chain V region (V3a) - human (fragment)  
S24211  
C:Species: Homo sapiens (man)  
C>Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999  
C:Accession: S24211  
R:Pargent, W.; Meindl, A.; Thiebe, R.; Mitzel, S.; Zachau, H.G.  
Eur. J. Immunol. 21, 1821-1827, 1991  
A:Title: The human immunoglobulin kappa locus. Characterization of the duplicated O r  
A:Reference number: S24205; MUID:91350953  
A:Accession: S24211  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-83 <PAR>  
A:Cross-references: EMBL:X59317; NID:q33270; PIDN:CAA42004.1; PID:q33271  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 33; DB 2; Length 83;  
Best Local Similarity 100.0%; Pred. No. 0.68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLKISR 7  
Db 59 FTLKISR 65

## RESULT 7

Ig kappa chain V region (patient 19 and 20) - human (fragment)  
S34091  
C:Species: Homo sapiens (man)  
C>Date: 13-Jan-1995 #sequence\_revision 13-Mar-1998 #text\_change 21-Jan-2000  
C:Accession: S34091  
R:Wagner, S.D.; Luzzatto, L.  
Eur. J. Immunol. 23, 391-397, 1993  
A:Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distrib  
A:Reference number: S34076; MUID:93170387  
A:Accession: S34091  
A:Molecule type: DNA  
A:Residues: 1-87 <NAG>  
A:Cross-references: EMBL:X67176  
A:Experimental source: patient 20  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:1-82/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 33; DB 2; Length 87;  
Best Local Similarity 100.0%; Pred. No. 0.72;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FTLKISR 7  
|||||||  
Db 63 FTLKISR 69

## RESULT 8

S34094  
Ig kappa chain V region - human  
C:Species: Homo sapiens (man)  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 21-Jan-2000  
C:Accession: S34094  
R:Wagner, S.D.; Luzzatto, L.  
R:Mo, J.A.; Bona, C.A.; Holmdahl, R.  
A:Title: V-kappa gene segments rearranged in chronic lymphocytic leukemia are distributed  
A:Reference number: S34076; MUID:93170387  
A:Accession: S34094  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-87 <MAG>  
A:Cross-references: EMBL:X67178  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F:3-82/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 33; DB 2; Length 87;  
Best Local Similarity 100.0%; Pred. No. 0.72;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FTLKISR 7  
|||||||  
Db 63 FTLKISR 69

## RESULT 9

S25155  
Ig kappa chain V region (SW1.5) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 28-Dec-1987 #sequence\_revision 02-Jun-1988 #text\_change 16-Aug-1996  
R:Accession: B25155  
R:Robbins, P.F.; Rosen, E.M.; Haba, S.; Nisonoff, A.  
Proc. Natl. Acad. Sci. U.S.A. 83, 1050-1054, 1986  
A:Title: Relationship of V(H) and V(L) genes encoding three idiotypic families of anti-F  
A:Reference number: A94083; MUID:86149212  
A:Accession: B25155  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-89 <ROB>  
C:Comment: This chain is expressed in an IGM with anti-arsenate activity.  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin

Query Match 100.0%; Score 33; DB 2; Length 89;  
Best Local Similarity 100.0%; Pred. No. 0.73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FTLKISR 7  
|||||||  
Db 54 FTLKISR 60

## RESULT 10

S42186  
Ig kappa chain V region - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 28-Sep-1994 #sequence\_revision 21-Jul-1995 #text\_change 21-Jan-2000  
C:Accession: S42186; S42195  
R:Mo, J.A.; Bona, C.A.; Holmdahl, R.  
Eur. J. Immunol. 23, 2503-2510, 1993  
A:Title: Variable region gene selection of immunoglobulin G-expressing B cells with spec  
A:Reference number: S42176; MUID:94009207  
A:Accession: S42186

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-91 <MOJ>  
A:Cross-references: EMBL:Z25442; NID:9407830; PTDN:CAA80929.1; PID:9407831  
A:Note: the authors translated the codon GTT for residue 81 as Thr, TGC for residue 8

A:Accession: S42195  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-91 <MO2>  
A:Cross-references: EMBL:Z25460; NID:9407848; PTDN:CAA80947.1; PID:9407849  
A:Note: the authors translated the codon GTT for residue 81 as Thr, TGC for residue 8  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F:7-86/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 33; DB 2; Length 91;  
Best Local Similarity 100.0%; Pred. No. 0.75;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FTLKISR 7  
|||||||  
Db 67 FTLKISR 73

## RESULT 11

S25463  
Ig kappa chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999  
R:Monestier, M.; Fasy, T.M.; Losman, M.J.; Novick, K.E.; Muller, S.  
submitted to the EMBL Data Library, July 1992  
A:Description: Structure and binding properties of monoclonal antibodies to core hist  
A:Reference number: S25174  
A:Accession: S25463  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-91 <MON>  
A:Cross-references: EMBL:X67624; NID:952189; PTDN:CAA47882.1; PID:9938264  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin

Query Match 100.0%; Score 33; DB 2; Length 91;  
Best Local Similarity 100.0%; Pred. No. 0.75;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FTLKISR 7  
|||||||  
Db 76 FTLKISR 82

## RESULT 12

PH1039  
Ig light chain V region (clone 165.27) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 15-Jun-1996  
C:Accession: PH1039  
R:Tillman, D.M.; Jou, N.T.; Hill, R.J.; Marion, T.N.  
J. Exp. Med. 176, 761-779, 1992  
A:Title: Both Igm and Igg anti-DNA antibodies are the products of clonally selective  
A:Reference number: PH0971; MUID:92381444  
A:Accession: PH1039  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-93 <TIL>  
A:Experimental source: B cell, strain [NZB x NZW]F1  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: immunoglobulin

Query Match 100.0%; Score 33; DB 2; Length 93;

Best Local Similarity 100.0%; Pred. No. 0.77;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLKISR 7  
|||||  
Db 67 FTLKISR 73

## RESULT 13

PL0258  
Ig kappa chain V region (anti-DNA, IE81VK) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 16-Aug-1996  
C:Accession: PL0258  
R:Shlomchik, M.; Mascelli, M.; Shan, H.; Radic, M.Z.; Pisetsky, D.; Marshak-Rothstein, A.  
J. Exp. Med. 171, 265-297, 1990  
A:Title: Anti-DNA antibodies from autoimmune mice arise by clonal expansion and somatic  
A:Reference number: PL0231; MUID:9011018  
A:Accession: PL0258  
A:Molecule type: mRNA  
A:Residues: 1-94 <SHL>  
C:Superfamily: Immunoglobulin V region: immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:1-6/Region: framework 1  
F:7-22/Region: complementarity-determining 1  
F:23-37/Region: framework 2  
F:38-44/Region: complementarity-determining 2  
F:45-76/Region: framework 3  
F:77-85/Region: complementarity-determining 3  
F:86-94/Region: framework 4

Query Match 100.0%; Score 33; DB 2; Length 94;

Best Local Similarity 100.0%; Pred. No. 0.77;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLKISR 7  
|||||  
Db 59 FTLKISR 65

## RESULT 14

S40320  
Ig kappa chain - human  
C:Species: Homo sapiens (man)  
C:Date: 06-Mar-1994 #sequence\_revision 26-May-1995 #text\_change 23-Jul-1999  
C:Accession: S40320  
R:Klein, R.; Jaenichen, R.; Zachau, H.G.  
Eur. J. Immunol. 23, 3248-3271, 1993  
A:Title: Expressed human immunoglobulin chi genes and their hypermutation.  
A:Reference number: S40312; MUID:94080891  
A:Accession: S40320  
A>Status: preliminary; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-96 <KLE>  
A:Cross-references: EMBL:X72430; NID:q441328; PIDN:CAAS1098.1; PID:q441329  
C:Superfamily: Immunoglobulin V region: immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 33; DB 2; Length 96;

Best Local Similarity 100.0%; Pred. No. 0.79;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLKISR 7  
|||||  
Db 59 FTLKISR 65

## RESULT 15

S24501  
Ig kappa chain V region - mouse  
C:Species: Mus musculus (house mouse)

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 21-Jan-2000  
C:Accession: S24501

R:Kaartinen, M.  
submitted to the EMBL Data Library, October 1991

A:Reference number: S24490  
A:Accession: S24501

A>Status: preliminary  
A:Molecule type: mRNA

A:Residues: 1-99 <RNA>

A:Cross-references: EMBL:X6638  
C:Superfamily: Immunoglobulin V region: immunoglobulin homology

C:Keywords: heterotetramer; immunoglobulin  
F:16-95/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 33; DB 2; Length 99;

Best Local Similarity 100.0%; Pred. No. 0.81;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLKISR 7  
|||||  
Db 76 FTLKISR 82

Search completed: July 15, 2002, 13:01:01  
Job time: 467 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: July 15, 2002, 13:23:32 ; Search time 44.9 Seconds  
(without alignments)  
6.036 Million cell updates/sec

Title:	US-09-712-819A-6
Perfect score:	33
Sequence:	1 FTLKISR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

```
Searched:      105224 segs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```

post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

Database : SwissProt\_40:\*

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	33	100.0	112	1	KV2C_HUMAN	P01616 homo sapien
2	33	100.0	113	1	KV2D_HUMAN	P01617 homo sapien
3	33	100.0	113	1	KV2G_MOUSE	P01631 mus musculu
4	33	100.0	115	1	KV2A_HUMAN	P01614 homo sapien
5	33	100.0	117	1	KV2E_HUMAN	P06309 homo sapien
6	33	100.0	133	1	KV2F_HUMAN	P06310 homo sapien
7	30	90.9	112	1	KV2D_MOUSE	P01629 mus musculu
8	30	90.9	113	1	KV2B_HUMAN	P01615 homo sapien
9	30	90.9	113	1	KV2E_MOUSE	P03976 mus musculu
10	30	90.9	113	1	KV2F_MOUSE	P01630 mus musculu
11	30	90.9	512	1	AMN3_HUMAN	P06078 homo sapien
12	30	90.9	772	1	C1PB_CLOTM	P01866 clostridium
13	30	90.9	1853	1	C1PA_CLOTM	P06851 clostridium
14	29	87.9	112	1	KV2A_MOUSE	P01628 mus musculu
15	29	87.9	113	1	KV2C_MOUSE	P01628 mus musculu
16	29	87.9	120	1	KV2B_MOUSE	P01627 mus musculu
17	29	87.9	834	1	PL5B_PSEAE	O91947 pseudomonas
18	28	84.8	259	1	K068_CHTRU	O84071 chlamydia t
19	28	84.8	259	1	K339_CHTRU	O90911 chlamydia r
20	28	84.8	711	1	CAD2_LISMO	P06004 listeria mc
21	28	84.8	745	1	KATB_ARATH	P40884 arabidopsis
22	27	81.8	100	1	KV3C_HUMAN	P01621 homo sapien
23	27	81.8	108	1	KV3A_HUMAN	P01619 homo sapien
24	27	81.8	109	1	KV3B_HUMAN	P01620 homo sapien
25	27	81.8	109	1	KV3D_HUMAN	P01622 homo sapien
26	27	81.8	109	1	KV3E_HUMAN	P01623 homo sapien
27	27	81.8	127	1	KV3G_HUMAN	P04206 homo sapien
28	27	81.8	127	1	KJ07_YEAST	P47105 saccharomye
29	27	81.8	129	1	KV3H_HUMAN	P04207 homo sapien
30	27	81.8	129	1	KV3I_HUMAN	P01815 homo sapien
31	27	81.8	129	1	KV3K_HUMAN	P01830 homo sapien
32	27	81.8	295	1	T008_YEAST	P38805 saccharomye
33	27	81.8	295	1	ATNC_BUFEA	P43002 bufo marinu

34	27	81.8	3.52	1	RECFE_PROMI	F22839	proteum	mr
35	27	81.8	1.231	1	KP4A_MOUSE	P33143	mus	musculu
36	27	81.8	1.232	1	KP4A_HUMAN	O95429	homo	sapien
37	27	81.8	1.541	1	NK3A_HUMAN	O89450	homo	sapien
38	27	81.8	1.578	1	NK3A_RAT	O07310	rattus	nory
39	27	81.8	2.068	1	T2D1_DROME	P51123	drosophila	
40	26	78.8	1.08	1	KV1_CANFA	P01618	canis	familia
41	26	78.8	1.72	1	R1M0_STRYP	P58183	streptococc	
42	26	78.8	1.14	1	V4H0_HRISP	P50361	trichobium	s
43	26	78.8	3.84	1	V984_METTH	O27075	methanobact	
44	26	78.8	4.29	1	YPR4_ECOLI	P76470	escherichia	
45	26	78.8	670	1	YB12_SCHPO	Q10139	schizosacch	

## ALIGNMENTS

RESULT	1			
ID	KV2C_HUMAN	STANDARD;	PRT;	112 AA.
CD	P01616;			
DT	21-JUL-1986 (rel. 01, Created)			
DT	21-JUL-1986 (rel. 01, Last sequence update)			
DT	15-JUL-1999 (rel. 38, Last annotation update)			
DE	Ig kappa chain V-II region ML.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE.			
RA	Dreyer W.J., Gray W.R., Hood L.E.;			
RT	"The genetic, molecular, and cellular basis of antibody formation:			
RT	some facts and a unifying hypothesis."			
RL	Cold Spring Harb. Symp. Quant. Biol. 32:353-367(1967).			
CC	1- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (3) MARKER			
CC	1- MISCELLANEOUS: THIS IS A BENICE-JONES PROTEIN.			
DR	PIR: A01877; K2HMUL.			
DR	InterPro: IPR0031006; IG_MHC.			
DR	InterPro: IPR003596; IG_V.			
DR	Pfam: PF00047; IgV_1.			
DR	SMART: SM00406; IgV_1.			
RW	Immunoglobulin V region; Benice-Jones protein.			
FT	DOMAIN 1 23			
FT	DOMAIN 24 38			
FT	DOMAIN 39 53			
FT	DOMAIN 54 60			
FT	DOMAIN 61 92			
FT	DOMAIN 93 101			
FT	DOMAIN 102 111			
FT	DISULFID 123 92			
FT	NON_TER 112 112			
QO	SEQUENCE 112 AA; 12055 MW; E5B222EFA7A4BE481 CRC64;			

Query Match	100.0%;	Score 33;	DB 1;	Length 112;
Best Local Similarity	100.0%;	Pred. No. 0.33;		
Matches	7;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;

```

      VY          L FTLKISR /
                |||||
Db           75   FTLKISR 81

RESULT_ 2
KV2D_HUMAN
ID       KV2D_HUMAN        STANDARD;    PRT;     113 AA.
AC       P01617;
DT       21-JUL-1986 (rel. 01, Created)
DI       21-JUL-1986 (rel. 01, Last sequence update)
DE       15-JUL-1999 (rel. 38, Last annotation update).
DS       Ig kappa chain V-II region TEW.
OS       Homo sapiens (Human).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxId=9606;  
[1]  
RP SEQUENCE (BENCE-JONES PROTEIN TEM).  
RA MEDLINE=74148480; PubMed=4596149;  
RA Putnam F.W., Whitley E.J., Jr., Paul C., Davidson J.N.;  
RT "Amino acid sequence of a kappa Bence Jones protein from a case of  
RT primary amyloidosis.";  
RN Biochemistry 12:3763-3780(1973).  
[2]  
RP SEQUENCE OF 1-27 (AMYLOID PROTEIN TEM).  
RA MEDLINE=73166638; PubMed=4700485;  
RA Terry W.D., Page D.L., Kimura S., Isobe T., Osseman E.F.,  
RA Glenner G.G.;  
RT "Structural identity of Bence Jones and amyloid fibril proteins in a  
RT patient with plasma cell dyscrasia and amyloidosis.";  
RN J. Clin. Invest. 52:1276-1281(1973).  
CC -1- MISCELLANEOUS: THE MAJOR AMYLOID PROTEIN APPEARS TO BE IDENTICAL  
CC WITH THE BENCE JONES PROTEIN ISOLATED FROM THE URINE OF A  
CC PATIENT WITH PLASMA CELL DYSKRASIA AND AMYLOIDOSIS.  
CC -1- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (1,2)  
CC MARKER.  
CC PIR: A01888; K2HUTW.  
DR InterPro: IPR003006; I9\_MHC.  
DR InterPro: IPR003596; I9\_v.  
DR Pfam: PF00047; I9\_1.  
DR SMART: SM00406; IGV\_1.  
KM Immunoglobulin V region; Bence-Jones protein; Amyloid.  
FT DOMAIN 1 23  
FT DOMAIN 24 39  
FT DOMAIN 40 54  
FT DOMAIN 55 61  
FT DOMAIN 62 93  
FT DOMAIN 94 102  
FT DOMAIN 103 112  
FT DISULFID 23 93  
FT NON\_TER 113  
SQ SEQUENCE 113 AA; 12316 MW; 0C3C38F81F1843CA CRC64;  
  
Query Match 100.0%; Score 33; DB 1; Length 113;  
Best Local Similarity 100.0%; Pred. No. 0.33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 FTLKISR 7  
Db 76 FTLKISR 82  
[1]  
[2]  
RESULT 3  
ID KV2G\_MOUSE STANDARD; PRT; 113 AA.  
AC P01631;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Ig kappa chain V-II region 26-10.  
OS Mus musculus (mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxId=10050;  
RN [1]  
RP SEQUENCE.  
RC STRAIN=A/J;  
RA MEDLINE=63178921; PubMed=6404298;  
RA Novotny J., Margolles M.N.;  
RT "Amino acid sequence of the light chain variable region from a mouse  
RT anti-digoxin hybridoma antibody.";  
RN Biochemistry 22:1153-1158(1983).  
CC -1- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM AN IGG2A HYBRIDOMA  
CC PROTEIN THAT BINDS DIGOXIN.

DR PIR: A01914; KVM526.  
DR InterPro: IPR003006; I9\_MHC.  
DR InterPro: IPR003596; I9\_v.  
DR Pfam: PF00047; I9\_1.  
DR SMART: SM00406; IGV\_1.  
KM Immunoglobulin V region; Monoclonal antibody; Hybridoma.  
FT DOMAIN 1 23  
FT DOMAIN 24 39  
FT DOMAIN 40 54  
FT DOMAIN 55 61  
FT DOMAIN 62 93  
FT DOMAIN 94 102  
FT DOMAIN 103 112  
FT DISULFID 23 93  
FT NON\_TER 113  
SQ SEQUENCE 113 AA; 12273 MW; F9F39CE949A84C2A CRC64;

Query Match 100.0%; Score 33; DB 1; Length 113;  
Best Local Similarity 100.0%; Pred. No. 0.33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLKISR 7  
Db 76 FTLKISR 82  
[1]  
[2]

RESULT 4  
ID KV2A\_HUMAN STANDARD; PRT; 115 AA.  
AC P01614;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Ig kappa chain V-II region Cmu.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=68242259; PubMed=5586923;  
RA Hilschmann N.;  
RT "The complete amino acid sequence of Bence Jones protein Cmu (kappa-  
RT type).";  
RN Hoppe-Seyler's Z. Physiol. Chem. 348:1718-1722(1967).  
RN REVISIONS TO 50; 52; 96 AND 97.  
RX MEDLINE=70063440; PubMed=4188189;  
RA Hilschmann N.;  
RT "Molecular basis of antibody formation.";  
RN Naturwissenschaften 56:195-205(1969).  
CC -1- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (3) MARKER.  
CC -1- MISCELLANEOUS: THIS IS A BENCE-JONES PROTEIN.  
CC PIR: A01885; K2HUCM.  
DR InterPro: IPR003006; I9\_MHC.  
DR InterPro: IPR003596; I9\_v.  
DR Pfam: PF00047; I9\_1.  
DR SMART: SM00406; IGV\_1.  
KM Immunoglobulin V region; Bence-Jones protein.  
FT DISULFID 24 95  
FT NON\_TER 115  
SQ SEQUENCE 115 AA; 12676 MW; 59E9P90A379569EC CRC64;

Query Match 100.0%; Score 33; DB 1; Length 115;  
Best Local Similarity 100.0%; Pred. No. 0.34;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLKISR 7  
Db 78 FTLKISR 84  
[1]  
[2]



```

RESULT 5
KV2E_HUMAN STANDARD; PRT; 117 AA.
AC P06309;
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DE 01-JAN-1988 (Rel. 06, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-II region GK607 precursor (Fragment).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84191506; PubMed=6325927;
RA Klobbeck H.G., Solomon A., Zachau H.G.;
RT "Contribution of human V kappa II germ-line genes to light-chain
diversity."
RL Nature 309:73-76(1984).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
CC EMBL; Z00009; -; NOT_ANNOTATED_CDS.
DR PIR; A01889; K2HUGM.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR Immunoglobulin V region; Signal.
FT NON_TER 1 1
FT SIGNAL <1 4
FT CHAIN 5 117 IG KAPPA CHAIN V-II REGION GK607.
FT DOMAIN 28 43 FRAMEWORK-1.
FT DOMAIN 44 58 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 59 65 FRAMEWORK-2.
FT DOMAIN 66 97 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 98 106 FRAMEWORK-3.
FT DOMAIN 107 116 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 107 116 FRAMEWORK-4.
FT NON_TER 117 117 BY SIMILARITY.
SO SEQUENCE 117 AA; 12664 MW; 92C57DC719E558B1 CRC64;

Query Match 100.0%; Score 33; DB 1; Length 117;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLKISR 7
DB 80 FTLKISR 86

RESULT 6
KV2E_HUMAN STANDARD; PRT; 133 AA.
AC P06310;
DT 01-JAN-1988 (Rel. 06, Created)
DE 01-JAN-1988 (Rel. 06, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-II region RPM1 6410 precursor.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=86041852; PubMed=2997711;
RA Klobbeck H.G., Meindl A., Combrat G., Zachau H.G.;
RT "Human immunoglobulin kappa light chain genes of subgroups II and
III."
RL Nucleic Acids Res. 13:6499-6513(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
CC EMBL; Z00020; CAA77315.1; -
DR PIR; A01890; K2HURP.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
DR Immunoglobulin V region; Signal.
FT SIGNAL 1 20
FT CHAIN 21 133 IG KAPPA CHAIN V-II REGION RPM1 6410.
FT DOMAIN 21 43 FRAMEWORK-1.
FT DOMAIN 44 59 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 60 74 FRAMEWORK-2.
FT DOMAIN 75 81 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 82 113 FRAMEWORK-3.
FT DOMAIN 114 122 FRAMEWORK-4.
FT DOMAIN 123 132 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 123 132 FRAMEWORK-4.
FT NON_TER 133 133 BY SIMILARITY.
SO SEQUENCE 133 AA; 14707 MW; 513C6AF3673009EE CRC64;

Query Match 100.0%; Score 33; DB 1; Length 133;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FTLKISR 7
DB 96 FTLKISR 102

RESULT 7
KV2D_MOUSE STANDARD; PRT; 112 AA.
AC P01629;
DT 21-JUL-1986 (Rel. 01, Created)
DE 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-II region 2S1.3.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=83055101; PubMed=7141411;
RA Herbst H., Chang J.Y., Aebersold R., Braun D.G.;
RT "Murine VK25 isotype sequence: monoclonal antibody 2S1.3 specific for
the group A streptococcal polysaccharide."
RL Hoppe-Seyler's Z. Physiol. Chem. 363:1069-1076(1982).
CC -I- MISCCELLANEOUS; THIS CHAIN IS FROM A HYBRIDOMA-DERIVED MONOCLONAL
ANTIBODY AGAINST THE STREPTOCOCCAL GROUP A POLYSACCHARIDE.
CC PIR; A01911; KWSST1.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
KW Immunoglobulin V region; Monoclonal antibody.
FT DOMAIN 1 23 FRAMEWORK-1.

```

```

FT DOMAIN 24 39 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 40 54 FRAMEWORK-2.
FT DOMAIN 55 61 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 62 93 FRAMEWORK-3.
FT DOMAIN 94 102 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 103 112 FRAMEWORK-4.
FT DISULFID 23 93 BY SIMILARITY.
FT NON_TER 112 112
SQ SEQUENCE 112 AA; 12221 MW; BD5F5E6D789FBBC CRC64;

```

```

Query Match 90.9%; Score 30; DB 1; Length 112;
Best Local Similarity 85.7%; Pred. No. 1.8;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 FTLKISR 7
   |||||
DB 76 FTLKISR 82

```

```

RESULT 8
KV2B_HUMAN STANDARD; PRT; 113 AA.
ID P01615;
AC 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-II region FR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=76253627; PubMed=821524;
RA Riesen W.F., Jaton J.-C.;
RT "Variable region sequence of the light chain from a Waldenström's
RL B-lymphoma with specificity for phosphorylcholine."
CC 1-1 MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A WALDENSTROM'S
CC MACROGLOBULIN THAT BINDS PHOSPHORYLCHOLINE.
DR PIR; A01886; K2HUPR.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
KW Immunoglobulin V region.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 39 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 40 54 FRAMEWORK-2.
FT DOMAIN 55 61 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 62 93 FRAMEWORK-3.
FT DOMAIN 94 102 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 103 112 FRAMEWORK-4.
FT DISULFID 23 93 BY SIMILARITY.
FT NON_TER 113 113
SQ SEQUENCE 113 AA; 12660 MW; 0C0DA39E46DB96DE CRC64;

```

```

Query Match 90.9%; Score 30; DB 1; Length 113;
Best Local Similarity 85.7%; Pred. No. 1.8;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 FTLKISR 7
   |||||
DB 76 FTLKISR 82

```

```

RESULT 9
KV2E_MOUSE STANDARD; PRT; 113 AA.
ID P03976;
AC 23-OCT-1986 (Rel. 02, Created)

```

```

DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-II region 17529.1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RC TISSUE=Hybridoma;
RX MEDLINE=85128968; PubMed=6441768;
RA Aebbersold R., Herbst H., Gutter T., Chang J.Y., Braun D.G.;
RT "Murine V kappa 25 and V kappa 27 amino-acid sequences of C57Bl/6
RT origin: monoclonal antibodies 17529.1 and 22525.1 specific for the
RT group A-streptococcal polysaccharide."
RL Hoppe-Seyler's Z. Physiol. Chem. 365:1375-1383(1984).
CC 1-1 MISCELLANEOUS: ANTI-STREPTOCOCCAL GROUP A CARBOHYDRATE ANTIBODY.
DR PIR; A01912; KVM517.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
KW Immunoglobulin V region; Hybridoma.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 39 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 40 54 FRAMEWORK-2.
FT DOMAIN 55 61 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 62 93 FRAMEWORK-3.
FT DOMAIN 94 102 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 103 112 FRAMEWORK-4.
FT DISULFID 23 93 BY SIMILARITY.
FT NON_TER 113 113
SQ SEQUENCE 113 AA; 12390 MW; 4E93797046F8DB33 CRC64;

```

```

Query Match 90.9%; Score 30; DB 1; Length 113;
Best Local Similarity 85.7%; Pred. No. 1.8;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 FTLKISR 7
   |||||
DB 76 FTLKISR 82

```

```

RESULT 10
KV2F_MOUSE STANDARD; PRT; 113 AA.
ID P01630;
AC 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-II region 7534.1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=83256427; PubMed=6409088;
RA Chang J.-Y., Herbst H., Aebbersold R., Braun D.G.;
RT "A new isotype sequence (V kappa 27) of the variable region of kappa-
RT light chains from a mouse hybridoma-derived anti-streptococcal group
RT A polysaccharide) antibody containing an additional cysteine residue.
RT Application of the dimethylaminobenzene isothiocyanate technique
RT for the isolation of peptides."
RL Biochem. J. 211:173-180(1983).
CC 1-1 MISCELLANEOUS: THIS CHAIN IS FROM A HYBRIDOMA-DERIVED MONOCLONAL
CC ANTIBODY AGAINST THE STREPTOCOCCAL GROUP A POLYSACCHARIDE.
DR PIR; A01913; KVM57S.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.

```

```
KW Immunoglobulin V region; Monoclonal antibody; Hybridoma.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 39 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 40 54 FRAMEWORK-2.
FT DOMAIN 55 61 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 62 93 FRAMEWORK-3.
FT DOMAIN 94 102 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 103 112 FRAMEWORK-4.
FT DISULFID 23 93 BY SIMILARITY.
FT NON_TER 113 113
SQ SEQUENCE 113 AA; 12496 MW; 42C019D10ADA3C91 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 113;
Best Local Similarity 85.7%; Pred. No. 1.8;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLKISR 7
Db 76 FTLKISR 82

RESULT 11
ANM3_HUMAN STANDARD; PRT: 512 AA.
AC 060678;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein arginine N-methyltransferase 3 (EC 2.1.1.-) (Fragment).
GN PRMT3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98307932; PubMed=9642256;
RA Tang J., Gary J.D., Clarke S., Herschman H.R.;
RT "PRMT 3, a type I protein arginine N-methyltransferase that differs
RT from PRMT 1 in its oligomerization, subcellular localization,
RT substrate specificity, and regulation."
RL J. Biol. Chem. 273:16935-16945(1998).
CC -1- FUNCTION: METHYLATES (MONO AND ASYMMETRIC DIMETHYLATION) THE
CC GUANIDINO NITROGENS OF ARGINYL RESIDUES IN SOME PROTEINS.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE PROTEIN ARGININE N-METHYLTRANSFERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF059531; AAC39937.1; -.
CC DR MIN; 603190; -.
CC DR InterPro: IPR000051; SAM_bind.
CC DR InterPro: IPR000822; ZnF-C2H2.
CC DR SMART: SM00355; ZnF_C2H2.1.
CC DR PROSITE: PS00028; ZINC_FINGER_C2H2.1; UNKNOWN_1.
CC KW Transferase; Methyltransferase; Zinc-finger.
CC FT NON_TER 1 1
CC FT ZN_FING 29 52 C2H2-TYPE.
CC FT SEQUENCE 512 AA; 58098 MW; 26DC4EB25162A18 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 512;
Best Local Similarity 85.7%; Pred. No. 8.8;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 FTLKISR 7
Db 419 FTLKISR 425

RESULT 12
CIPB_CLOTH STANDARD; PRT: 772 AA.
ID CIPB_CLOTH
AC Q01866;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein SI/SI)
DE (Cellulose Integrating protein B) (Fragment).
GN CIPB.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS;
RX MEDLINE=93146373; PubMed=1490597;
RA Poole D.M., Morag E., Jamed R., Bayer E.A., Hazlewood G.P.,
RA Gilbert H.U.;
RT "Identification of the cellulose-binding domain of the cellulosome
RT subunit SI from Clostridium thermocellum YS."
RL FEMS Microbiol. Lett. 78:181-186(1992).
CC -1- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
CC CELLULYTIC ENZYMES PROBABLY THROUGH THE BINDING OF THE NINE
CC REPEATED DOMAINS WITH THE DUPLICATED SEGMENT OF 24 AMINO ACIDS
CC PRESENT IN CATALYTIC SUBUNITS OF THE CELLULOSE.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE.
CC -1- SIMILARITY: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X68233; CAA48312.1; -.
CC DR HSSP: Q06851; IMBC.
CC DR InterPro: IPR001956; CBD_3.
CC DR InterPro: IPR002102; Cohesin.
CC DR InterPro: IPR002105; Dockerin_1.
CC DR InterPro: IPR002048; EF-hand.
CC DR Pfam: PF00942; CBD_3; 1.
CC DR Pfam: PF00953; Cohesin; 3.
CC DR Pfam: PF00404; Dockerin_1; 2.
CC DR PRODOM: PD001947; CBD_3; 1.
CC DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
CC DR PROSITE: PS00448; CLOS_CELLULOSE_RPT; 2.
CC KW Cellulose degradation; Cell wall; Glycoprotein; Repeat.
CC FT NON_TER 1 1
CC FT DOMAIN <1 607 3 X 147 AA APPROXIMATE REPEATS.
CC FT REPEAT <1 80 1 (PARTIAL).
CC FT REPEAT 94 240 2.
CC FT REPEAT 462 607 3.
CC FT DOMAIN 81 93 PRO/THR-RICH.
CC FT DOMAIN 241 272 PRO/THR-RICH.
CC FT DOMAIN 273 439 CELLULOSE-BINDING.
CC FT DOMAIN 440 461 PRO/THR-RICH.
CC FT DOMAIN 710 766 2 X 24 AA APPROXIMATE REPEATS.
CC FT REPEAT 710 733 2-1.
CC FT REPEAT 743 766 2-2.
CC FT SEQUENCE 772 AA; 82491 MW; BBF06DE5B094FE10 CRC64;
```

Query Match 90.9%; Score 30; DB 1; Length 772;  
 Best Local Similarity 85.7%; Pred No. 13;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLKISR 7  
 Db 672 YTLKISR 678

RESULT 13  
 CIPA\_CLOTH STANDARD; PRT: 1853 AA.

AC Q06851;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Cellulosomal scaffolding protein A precursor (Cellulosomal  
 glycoprotein st/sl) (Cellulose Integrating protein A) (Cohesin).  
 GN CIPA.  
 OS Clostridium thermocellum.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 CC Clostridium;  
 CC NCBI\_TaxID=1515;  
 OX [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-40 AND 42-43.  
 RC STRAIN-ATCC 27405;  
 RX MEDLINE=93302508; PubMed=8316083;  
 RA Genogross U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S.,  
 RA Demain A.L.;  
 RT "Sequencing of a Clostridium thermocellum gene (cipa) encoding the  
 RT cellulosomal St-protein reveals an unusual degree of internal  
 RT homology.";  
 RL Mol. Microbiol. 8:325-334(1993).  
 RN [2]  
 RP SEQUENCE OF 1820-1853 FROM N.A.  
 RX MEDLINE=93209931; PubMed=8458932;  
 RA Fujino T., Beguin P., Aubert J.-P.;  
 RT "Organization of a Clostridium thermocellum gene cluster encoding the  
 RT cellulosomal scaffolding protein Cipa and a protein possibly involved  
 RT in attachment of the cellulosome to the cell surface.";  
 RL J. Bacteriol. 175:1891-1899(1993).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.  
 RX MEDLINE=9728934; PubMed=9083107;  
 RA Shlomo L.J.W., Bayer E.A., Morag E., Yaron S., Shoham Y.,  
 RA Prolov F.;  
 RT "A cohesin domain from Clostridium thermocellum: the crystal  
 RT structure provides new insights into cellulosome assembly.";  
 RL Structure 5:381-390(1997).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.  
 RX MEDLINE=97076134; PubMed=8918451;  
 RA Torno J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,  
 RA Steltz T.A.;  
 RT "Crystal structure of a bacterial family-III cellulose-binding  
 RT domain: a general mechanism for attachment to cellulose.";  
 RL EMBO J. 15:5739-5751(1996).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.  
 RX MEDLINE=98022914; PubMed=9402065;  
 RA Tavares G.A., Beguin P., Alzari P.M.;  
 RT "The crystal structure of a type I cohesin domain at 1.7-A  
 RT resolution.";  
 RL J. Mol. Biol. 273:701-713(1997).  
 CC -1- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT  
 CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE  
 CC CELLULOLYTIC ENZYMES.  
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE.  
 CC -1- DOMAIN: CONTAINS 9 COPIES OF A DOMAIN (COHESIN) OF ABOUT 150  
 CC RESIDUES THAT BINDS TO THE DOCKERIN DOMAIN BORN BY THE CATALYTIC  
 CC COMPONENTS OF THE CELLULOSE.  
 CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS  
 CC WELL, AS IN OTHER C.THERMOCCELLUM CELLULOSE ENZYMES. THIS DOMAIN

CC MAY FUNCTION AS THE BINDING LIGAND FOR THE ST COMPONENT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 DR EMBL: L08665; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL: X67506; CAA47840.1; -.  
 DR PIR: S36859; S36859.  
 DR PDB: 1AND; 23-JUL-97.  
 DR PDB: 1AOH; 08-JUL-98.  
 DR PDB: 1NRC; 26-SEP-97.  
 DR InterPro: IPR001956; CBD\_3.  
 DR InterPro: IPR002102; Cohesin.  
 DR InterPro: IPR002105; Dockerin\_1.  
 DR InterPro: IPR02048; Ef-hand.  
 DR Pfam: PF00942; CBD\_3; 1.  
 DR Pfam: PF00963; Cohesin; 9.  
 DR Pfam: PF00404; Dockerin\_1; 2.  
 DR ProDom: PD001947; CBD\_3; 1.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN\_1.  
 DR PROSITE: PS00448; CLOS\_CELLULOSE\_RPT; 2.  
 KM Cellulose degradation; Cell wall; Glycoprotein; Repeat; Signal;  
 KM 3D-structure.  
 FT SIGNAL 1 28  
 FT CHAIN 29 1853  
 FT DOMAIN 364 522  
 FT DOMAIN 54 1694  
 FT REPEAT 29 182  
 FT REPEAT 183 322  
 FT REPEAT 560 704  
 FT REPEAT 724 866  
 FT REPEAT 889 1031  
 FT REPEAT 1054 1196  
 FT REPEAT 1219 1361  
 FT REPEAT 1384 1526  
 FT REPEAT 1548 1690  
 FT DOMAIN 1791 1846  
 FT REPEAT 1791 1814  
 FT REPEAT 1824 1847  
 FT REPEAT 1834 1847  
 FT CONFLICT 1615 1615  
 SO SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABFC5372 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 1853;  
 Best Local Similarity 85.7%; Pred. No. 33;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLKISR 7  
 Db 1753 YTLKISR 1759

RESULT 14  
 KVZA\_MOUSE STANDARD; PRT: 112 AA.

AC P01626;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain V-II region MOPC 167.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 GN NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=79000273; PubMed=99160;  
 RA Rudikoff S., Potter M.;

RT "kappa Chain variable region from M167, a phosphorylcholine binding

RT myeloma protein."

RL Biochemistry 17:2703-2707(1978).

CC -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT

CC BINDS PHOSPHORYLCHOLINE. THE SEQUENCE OF THE V REGION OF THE HEAVY

CC CHAIN HAS ALSO BEEN DETERMINED.

CC PIR: A01908: KYMS16.

DR InterPro: IPR003006; Ig\_MHC.

DR InterPro: IPR003596; Ig\_V.

DR Pfam: PF00047; Ig; 1.

DR SMART: SM00406; IgV; 1.

KW Immunoglobulin V region.

FT DOMAIN 1 23

FT DOMAIN 24 39

FT DOMAIN 40 54

FT DOMAIN 55 61

FT DOMAIN 62 93

FT DOMAIN 94 102

FT DOMAIN 103 112

FT DISULFID 23 93

FT NON\_TER 112 112

SO SEQUENCE 112 AA; 12349 MW; A58EDFD6404B9726 CRC64;

QY 1 FTUKISR 7

Db 76 FTLEISR 82

RESULT 15

KV2C\_MOUSE STANDARD; PRT: 113 AA.

AC P01628;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Ig kappa chain V-II region M09C 511.

OS Mus musculus (mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE.

RX MEDLINE=81052016; PubMed=6776396;

RA Appella E.;

RT "Amino acid sequence of the light chain variable region of M511, a

RT phosphorylcholine-binding murine myeloma protein.";

RL Mol. Immunol. 17:711-718(1980).

CC -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM A MYELOMA PROTEIN THAT

CC BINDS PHOSPHORYLCHOLINE.

CC PIR: A01910: KYMS1.

DR InterPro: IPR003006; Ig\_MHC.

DR InterPro: IPR003596; Ig\_V.

DR Pfam: PF00047; Ig; 1.

DR SMART: SM00406; IgV; 1.

KW Immunoglobulin V region.

FT DOMAIN 1 23

FT DOMAIN 24 39

FT DOMAIN 40 54

FT DOMAIN 55 61

FT DOMAIN 62 93

FT DOMAIN 94 102

FT DOMAIN 103 112

FT DISULFID 23 93

FT NON\_TER 113 113

SO SEQUENCE 113 AA; 12496 MW; EFB0DC4DA2BD3450 CRC64;

QY 1 FTUKISR 7

Db 76 FTLEISR 82

Query Match

Best Local Similarity 87.9%; Score 29; DB 1; Length 112;

Best Local Similarity 85.7%; Pred. No. 3.2;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Query Match

Best Local Similarity 87.9%; Score 29; DB 1; Length 113;

Best Local Similarity 85.7%; Pred. No. 3.3;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Query Match

Best Local Similarity 87.9%; Score 29; DB 1; Length 113;

Best Local Similarity 85.7%; Pred. No. 3.3;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Query Match

Best Local Similarity 87.9%; Score 29; DB 1; Length 113;

Best Local Similarity 85.7%; Pred. No. 3.3;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Query Match

Best Local Similarity 87.9%; Score 29; DB 1; Length 113;

Best Local Similarity 85.7%; Pred. No. 3.3;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Query Match

Best Local Similarity 87.9%; Score 29; DB 1; Length 113;

Best Local Similarity 85.7%; Pred. No. 3.3;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Query Match

Best Local Similarity 87.9%; Score 29; DB 1; Length 113;



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: July 15, 2002, 13:22:45 ; Search time 172.49 Seconds  
(without alignments)  
7.020 Million cell updates/sec

Title: US-09-712-819A-6  
Perfect score: 33  
Sequence: 1 FTLKISR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

SPTREMBL\_19:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.rodent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriophage:\*  
17: sp.archaeophages:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	% Length	DB ID	Description
1	33	100.0	104	11 Q9JL82	Q9JL82 mus musculu
2	33	100.0	114	4 Q9JL80	Q9JL80 Homo sapien
3	33	100.0	238	11 Q9JL87	Q9JL87 mus musculu
4	30	90.9	471	5 Q9XU13	Q9XU13 caenorhabdi
5	30	90.9	638	2 Q46453	Q46453 clostridium
6	30	90.9	3848	5 Q76737	Q76737 dictyosteli
7	28	84.8	31	2 Q9R4X1	Q9R4X1 treponema d
8	28	84.8	80	2 Q9R4X1	Q9R4X1 corynebacte
9	28	84.8	80	2 Q9R4X1	Q9R4X1 brevbacter
10	28	84.8	118	16 Q9R4X1	Q9R4X1 staphylococ
11	28	84.8	285	11 Q9R4X1	Q9R4X1 deinococcus
12	28	84.8	310	11 Q9C048	Q9C048 mus musculu
13	28	84.8	327	5 Q95Y35	Q95Y35 caenorhabdi
14	28	84.8	346	11 Q9C016	Q9C016 mus musculu
15	28	84.8	347	10 Q93XF8	Q93XF8 zea mays (m
16	28	84.8	389	16 Q25171	Q25171 helicobacte

17	28	84.8	389	16 Q9ZK68	Q9ZK68 helicobacte
18	28	84.8	399	2 Q56257	Q56257 treponema d
19	28	84.8	449	5 Q9VC17	Q9VC17 drosophila
20	28	84.8	477	5 Q9T267	Q9T267 caenorhabdi
21	28	84.8	546	4 Q9NM40	Q9NM40 homo sapien
22	28	84.8	564	16 Q9Y9P8	Q9Y9P8 streptococc
23	28	84.8	733	16 Q926C7	Q926C7 rhizobium m
24	28	84.8	744	10 Q9T047	Q9T047 arabidopsis
25	28	84.8	943	2 Q52149	Q52149 synechococc
26	28	84.8	980	10 Q949G8	Q949G8 malus flori
27	28	84.8	1015	10 Q949G9	Q949G9 malus flori
28	28	84.8	1490	5 Q9TX96	Q9TX96 drosophila
29	27	81.8	51	17 Q973A6	Q973A6 sulfolobus
30	27	81.8	109	4 Q9UL86	Q9UL86 homo sapien
31	27	81.8	109	4 Q9UL78	Q9UL78 homo sapien
32	27	81.8	147	3 Q08915	Q08915 saccharomyc
33	27	81.8	171	10 Q9C8P6	Q9C8P6 arabidopsis
34	27	81.8	204	16 Q9KMI2	Q9KMI2 VIBRIO chol
35	27	81.8	219	16 Q9RT85	Q9RT85 delinococcus
36	27	81.8	294	16 P72830	P72830 synchocyst
37	27	81.8	315	10 Q9ZU65	Q9ZU65 arabidopsis
38	27	81.8	316	2 Q9ADW5	Q9ADW5 entlichia c
39	27	81.8	367	16 Q968H5	Q968H5 rhizobium l
40	27	81.8	389	17 Q57712	Q57712 pyrococcus
41	27	81.8	437	16 Q9C3X3	Q9C3X3 pasteurrella
42	27	81.8	443	2 Q32628	Q32628 haemophilus
43	27	81.8	443	2 Q32624	Q32624 haemophilus
44	27	81.8	488	12 Q67635	Q67635 gallid hept
45	27	81.8	540	17 Q971R4	Q971R4 sulfolobus

## ALIGNMENTS

RESULT 1  
Q9JL82 PRELIMINARY; PRT; 104 AA.  
ID Q9JL82  
AC Q9JL82;  
DT 01-OCT-2000 (TREMBLrel. 15, created)  
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE AMYT-MYOSIN IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION  
DE (FRAGMENT).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_Taxid-10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C;  
RX MEDLINE=20448942; PubMed=10992488;  
RA Maikael S., Liao L., Cunningham M.W., Diamond B.;  
RT "T-Cell-dependent antibody response to the dominant epitope of  
RT streptococcal polysaccharide, N-acetyl-glucosamine, is cross-reactive  
RT with cardiac myosin.";  
RL Infect. Immun. 68:5803-5808 (2000).  
DR EMBL; AF206024; AAF69322.1; .  
DR HSSP; P01607; IRET.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig\_1.  
DR SMART; SM00406; IGV; 1.  
FT NON\_TER 1  
FT NON\_TER 104  
SQ SSEQUENCE 104 AA; 11360 MW; SDA8B8P5FOALAE CRC64;

Query Match 100.0%; Score 33; DB 11; Length 104;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FTLKISR 7  
|||||||

Db 68 FTLKISR 74

RESULT 2

ID Q9UL80 PRELIMINARY; PRT: 114 AA.

AC Q9UL80; 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE MYOSIN-REACTIVE IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98277139; PubMed=9614934;

RA Wu X., Liu B., Van der Werwe P.L., Kalis N.N., Berney S.M., Young D.C.;

RT "Myosin-reactive autoantibodies in rheumatic carditis and normal fetus.";

RL Clin. Immunol. Immunopathol. 87:184-192(1998).

DR EMBL: AF035034; AAD56270.1; .

DR HSSP: P80362; 1WT.

DR InterPro: IPR003006; Ig\_MHC.

DR InterPro: IPR003596; Ig\_v.

DR Pfam: PF00047; Ig\_1.

DR SMART: SM00406; IGV; 1.

FT NON\_TER 1

FT TER 114

SO SEQUENCE 114 AA; 12775 MW; 070E31E210DICB01 CRC64;

Query Match 100.0%; Score 33; DB 4; Length 114;

Best Local Similarity 100.0%; Pred. No. 1.6;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLKISR 7

Db 76 FTLKISR 82

RESULT 3

ID Q99M37 PRELIMINARY; PRT: 238 AA.

AC Q99M37; 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE HYPOTHETICAL 26.3 KDA PROTEIN.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD, CROSS TISSUE.

RA Strausberg R.;

RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL: BC002035; AA02035.1; .

DR HSSP: P01679; 2FBU.

DR InterPro: IPR003599; Ig.

DR InterPro: IPR003597; Ig\_c1.

DR InterPro: IPR003600; Ig\_like.

DR InterPro: IPR003006; Ig\_MHC.

DR InterPro: IPR003596; Ig\_v.

DR Pfam: PF00047; Ig\_2.

DR SMART: SM00409; IGV; 2.

DR SMART: SM00407; IGV; 1.

DR SMART: SM00406; IGV; 1.

DR SMART: SM00410; IGVlike; 1.

DR PROSITE: PS00290; IGV\_MHC; UNKNOWN\_1.

KW Hypothetical protein

SO SEQUENCE 238 AA; 26344 MW; FB2B06A0B801330A CRC64;

OY 1 FTLKISR 7

Db 95 FTLKISR 101

Query Match 100.0%; Score 33; DB 11; Length 238;

Best Local Similarity 100.0%; Pred. No. 3.3;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLKISR 7

Db 95 FTLKISR 101

RESULT 4

ID Q9XU13 PRELIMINARY; PRT: 471 AA.

AC Q9XU13; 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE ZK39.1 PROTEIN.

GN ZK39.1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Kershaw J.K.;

RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for investigating biology.";

RL Science 282:2012-2018(1998).

DR EMBL: Z82093; CAB05017.1; .

SO SEQUENCE 471 AA; 53577 MW; CB6759DFE21FF959 CRC64;

Query Match 90.9%; Score 30; DB 5; Length 471;

Best Local Similarity 85.7%; Pred. No. 37;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FTLKISR 7

Db 460 FTLKISR 466

RESULT 5

ID Q46453 PRELIMINARY; PRT: 638 AA.

AC Q46453; 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE CIPA PROTEIN (FRAGMENT).

GN CIPA.

OS Clostridium thermocellum.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.

OX NCBI\_TaxID=1515;

RN [1]

RP SEQUENCE FROM N.A.

RA Fujino T., Beguin P., Aubert J.P.;

RT Cloning of a Clostridium thermocellum DNA fragment encoding polypeptides that bind the catalytic components of the cellulosome.";

RL FEMS Microbiol. Lett. 94:165-170(1992).

RN [2]

RP SEQUENCE FROM N.A.



RX MEDLINE=92394486; PubMed=1521765;  
RA Fujino T., Beguin P., Aubert J.P.:  
RT "Cloning of a Clostridium thermocellum DNA fragment encoding  
RT polypeptides that bind the catalytic components of the cellulosome."  
RL FEMS Microbiol. Lett. 73:165-170(1992).  
DR EMBL; X67406; CAA47806.1; -  
DR HSSP; Q06851; IAOH.  
DR InterPro: IPR002102; Cohesin.  
DR InterPro: IPR002105; Dockerin\_1.  
DR InterPro: IPR002048; EF-hand.  
DR Pfam; PF00404; Dockerin\_L; 2.  
DR PROSITE; PS00448; CLOS\_CELLULOSOME\_RPT; UNKNOWN\_2.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_1.  
FT NON\_TER 1  
SQ SEQUENCE 638 AA; 68255 MW; 355C17E50AF0CB6 CRC64;

Query Match 90.9%; Score 30; DB 2; Length 638;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 FTLKISR 7  
Db 538 YTLKISR 544

RESULT 6  
ID 076737 PRELIMINARY; PRT; 3848 AA.  
AC 076737;  
DT 01-NOV-1998 (TREMblrel. 08, Created)  
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE TIPC.  
GN TIPC.  
OS Dictyostelium discoideum (slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostellium.  
OX NCBI\_Taxid=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX4;  
RA MEDLINE=9331084; PubMed=10402673;  
RT Stege J.T., Laub M.T., Loomis W.F.:  
RT "tip genes act in parallel pathways of early Dictyostellium  
RT development."  
RL Dev. Genet. 25:64-77(1999).  
DR EMBL; AF079445; AAC31916.1; -  
SQ SEQUENCE 3848 AA; 440033 MW; 65B671566F7AE534 CRC64;

Query Match 90.9%; Score 30; DB 5; Length 3848;  
Best Local Similarity 85.7%; Pred. No. 3e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FTLKISR 7  
Db 886 FTLKISR 892

RESULT 7  
ID 09R4X1 PRELIMINARY; PRT; 31 AA.  
AC 09R4X1;  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
DE HEMOLYSIN (FRAGMENT).  
OS Treponema denticola.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_Taxid=158;  
RN [1]  
RP SEQUENCE.

RX MEDLINE=94376628; PubMed=8090078;  
RA Chu L., Holt S.C.:  
RT "Purification and characterization of a 45 kDa hemolysin from  
RT Treponema denticola ATCC 35404."  
RL Microb. Pathog. 16:197-212(1994).  
SQ SEQUENCE 31 AA; 3650 MW; 054650358CBF24C4 CRC64;

Query Match 84.8%; Score 28; DB 2; Length 31;  
Best Local Similarity 85.7%; Pred. No. 8;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FTLKISR 7  
Db 5 FTLKISR 11

RESULT 8  
ID 09FDR9 PRELIMINARY; PRT; 80 AA.  
AC 09FDR9;  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
DE ATP1 PROTEIN.  
GN ATP1.  
OS Corynebacterium glutamicum (Brevibacterium flavum).  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
OC Corynebacterium.  
OX NCBI\_Taxid=1718;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC13060;  
RA Sekine H., Yokota A., Tomita F.:  
RT "Nucleotide sequence of alp operon of Corynebacterium glutamicum."  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB046112; BAB08150.1; -  
SQ SEQUENCE 80 AA; 9308 MW; 23C0E33367F8DE8 CRC64;

Query Match 84.8%; Score 28; DB 2; Length 80;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FTLKISR 6  
Db 36 FTLKISR 41

RESULT 9  
ID 09FAA7 PRELIMINARY; PRT; 80 AA.  
AC 09FAA7;  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
DE ATP1 PROTEIN.  
GN ATP1.  
OS Brevibacterium flavum.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae.  
OX NCBI\_Taxid=92706;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC14067;  
RA Sekine H., Tomita F., Yokota A.:  
RT "Nucleotide sequence of alp operon of Brevibacterium flavum."  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB048368; BAB13353.1; -  
SQ SEQUENCE 80 AA; 9339 MW; 72CD0E33367E4B33 CRC64;

```

Query Match      84.8%; Score 28; DB 2; Length 80;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FTLKIS 6
    |||||
Db 36 FTLKIS 41

RESULT 10
O99V56 PRELIMINARY; PRT; 118 AA.
ID O99V56
AC O99V56;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DR 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN SA0890 (HYPOTHETICAL PROTEIN SAV1037).
GN SA0890 OR SAV1037.
OS Staphylococcus aureus (strain N315), and
OC Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879, 158878;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-S.aureus (strain N315), and S.aureus (strain Mu50);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama T., Baba T., Yuzawa H., Kobayashi T.,
RA Cul L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-O., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mutsaers U. Y., Takahashi N. K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogatawara N., Hayashi H., Hiratazu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus."
RL Lancet 357:1225-1240(2001).
DR EMBL: AP003132; BAB42135.1; -.
DR EMBL: AP003361; BAB57199.1; -.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 118 AA; 13229 MW; A895E0Z45E3BC710 CRC64;

Query Match      84.8%; Score 28; DB 16; Length 118;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FTLKIS 7
    |||||
Db 80 FTLKIS 86

RESULT 11
O9RMV8 PRELIMINARY; PRT; 285 AA.
ID O9RMV8
AC O9RMV8;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DR 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DR 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE HYPOTHETICAL 31.2 KDA PROTEIN.
GN DR0527.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI.
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dogson R.J., Hart D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utechtack T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,

```

```

RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RI."
RL Science 286:1571-1577(1999).
DR EMBL: AE001911; AAF10109.1; -.
DR TIGR: DR0527; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 285 AA; 31219 MW; 02DDDFAD0D9351E3 CRC64;

Query Match      84.8%; Score 28; DB 16; Length 285;
Best Local Similarity 71.4%; Pred. No. 73;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FTLKIS 7
    |||||
Db 25 FTLKIS 31

RESULT 12
O9CUD8 PRELIMINARY; PRT; 310 AA.
ID O9CUD8
AC O9CUD8;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DR 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE 1110021E09RIK PROTEIN (FRAGMENT).
GN 1110021E09RIK.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eularchia; Rodentia; Sclurognathii; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=MEDULLA OBLONGATA;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shiba K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izaawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann M., Gasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirini L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boileau D., Bolunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata K., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AK018126; BAB31082.1; -.
DR MGD: MGI:1914171; 1110021E09RIK.
FT NON_TER 310
SQ SEQUENCE 310 AA; 35234 MW; 017744FA3AD1EBBC CRC64;

Query Match      84.8%; Score 28; DB 11; Length 310;
Best Local Similarity 71.4%; Pred. No. 79;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FTLKIS 7
    |||||
Db 137 FTLKIS 143

RESULT 13

```